

# “Pretty strong” converse for the quantum capacity of degradable channels

Ciara Morgan and Andreas Winter

**Abstract**—We exhibit a possible road towards a strong converse for the quantum capacity of degradable channels. In particular, we show that all degradable channels obey what we call a “pretty strong” converse: When the code rate increases above the quantum capacity, the fidelity makes a discontinuous jump from 1 to at most  $\frac{1}{\sqrt{2}}$ , asymptotically. A similar result can be shown for the private (classical) capacity.

Furthermore, we can show that if the strong converse holds for symmetric channels (which have quantum capacity zero), then degradable channels obey the strong converse: The above-mentioned asymptotic jump of the fidelity at the quantum capacity is then from 1 down to 0.

**Index Terms**—quantum information, private classical information, channel coding, strong converse, smooth entropies, error-rate trade-off

## I. INTRODUCTION

Communication via noisy channels is one of the information processing tasks by which, following the fundamental work of Shannon [36], we have learned to quantify information and noise. One of the most important models considered from these early days of information theory is that of a discrete memoryless channel, for which Shannon gave his famous single-letter formula for the capacity (i.e., the maximum communication rate achievable by asymptotically error-free block coding).

The analogous model in quantum Shannon theory is the memoryless quantum channel  $\mathcal{N}^{\otimes n}$  (for asymptotically large integer  $n$ ), given by a completely positive and trace preserving (cptp) map  $\mathcal{N} : \mathcal{L}(A') \rightarrow \mathcal{L}(B)$ , with Hilbert spaces  $A'$  and  $B$  that we assume to be finite dimensional throughout this paper.

The quantum capacity  $Q(\mathcal{N})$  of  $\mathcal{N}$  is informally defined as the maximum rate at which quantum information can be transmitted asymptotically faithfully over that channel, when using it  $n \rightarrow \infty$  times.

CM is with the Institut für Theoretische Physik, Leibniz Universität Hannover, Appelstraße 2, D-30167 Hannover, Germany. Email: ciara.morgan@itp.uni-hannover.de

AW is with ICREA and Física Teòrica: Informació i Fenòmens Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain. He is currently on leave from the Department of Mathematics, University of Bristol, Bristol BS8 1TW, U.K. Email: andreas.winter@uab.cat

This work began when both authors were with CQT, National University of Singapore, 3 Science Drive 2, Singapore 117543.

CM is funded from the EU grant QFTCMPS and the cluster of excellence EXC 201 Quantum Engineering and Space-Time Research. AW acknowledges financial support by the European Commission (STREP “QCS” and Integrated Project “QESSENCE”), the ERC (Advanced Grant “IRQUAT”), a Royal Society Wolfson Merit Award and a Philip Leverhulme Prize. The Centre for Quantum Technologies is funded by the Singapore Ministry of Education and the National Research Foundation as part of the Research Centres of Excellence programme.

As for all channel capacity theorems, the quantum capacity theorem consists of a direct part and a converse. The direct part states that for rates below a certain threshold there exist codes with decoding error (quantified as a certain distance from noiseless transmission) tending to 0 in the number of channel uses. The converse states that if the rate lies above this threshold then the error does not go to 0 for any sequence of codes. To be precise, this is known as a *weak converse* and the threshold rate sometimes called *weak capacity*. A *strong converse* is the statement that for rates above the capacity the error converges to its maximum 1 as  $n \rightarrow \infty$ .

Strong converse theorems have been shown to hold for other types of information sent over memoryless quantum channels, including classical information encoded into product states [29], [50] and for general input states (i.e. allowing the possibility of entangled input signal states) over certain classes of quantum channels, by [25]. The strong converse holds also for entanglement-assisted classical communication over memoryless quantum channels, by the Quantum Reverse Shannon Theorem [4], [6]. Strong converses do not hold by default; certain quantum channels with memory have a weak capacity but fail the strong converse [13], [18].

The paper is structured as follows: In Section II we recall the definition of codes, error criteria and the quantum capacity. Then, in Section III we discuss the weak converse for the quantum capacity and the possibility of strong converses. In Section IV, we review the concept of degradable channels and the analysis of Devetak and Shor [16] of their quantum capacity. We will present the argument in a form that will aid in the subsequent finer analysis, proving a structural lemma on degradable channels along the way. Then in Section V, we state and prove our first main result (Theorem 2) strongly bounding the rate of channels with sufficiently small error. All necessary auxiliary results are stated in this section, however the proofs are relegated to the appendix. Subsequently, we prove an analogous rate bound for the private classical capacity (Theorem 14 in Section VI), and then show that a strong converse for all symmetric channels implies the strong converse for all degradable channels (Theorem 19 in Section VII). In Section VIII we discuss a semidefinite programming approach to deal with the symmetric channels. We conclude in Section IX with a brief discussion of what was achieved and highlight open problems.

On notation: In this paper,  $\log$  is always the binary logarithm, and  $\exp$  its inverse, the exponential function to base 2. The natural logarithm is denoted  $\ln x$ , the natural exponential function  $e^x$ .

## II. QUANTUM CHANNEL CAPACITY

For a given channel  $\mathcal{N} : \mathcal{L}(A') \rightarrow \mathcal{L}(B)$ , we consider encoding and decoding of quantum information, given by completely positive and trace preserving (cptp) maps

$$\begin{aligned}\mathcal{E} : \mathcal{L}(C) &\rightarrow \mathcal{L}(A'), \\ \mathcal{D} : \mathcal{L}(B) &\rightarrow \mathcal{L}(C),\end{aligned}$$

which together form a *quantum code*. The idea is that the information to be sent is subjected to the overall effective channel  $\mathcal{D} \circ \mathcal{N} \circ \mathcal{E} : \mathcal{L}(C) \rightarrow \mathcal{L}(C)$ . For a Hilbert space  $\mathcal{H}$ , we denote by

$$\begin{aligned}\mathcal{S}(\mathcal{H}) &= \{\rho \geq 0 \text{ s.t. } \text{Tr } \rho = 1\}, \\ \mathcal{S}_{\leq}(\mathcal{H}) &= \{\rho \geq 0 \text{ s.t. } \text{Tr } \rho \leq 1\},\end{aligned}$$

the set of states and sub-normalized densities, respectively.

There are many ways of defining mathematically the notion that the output is a good approximation of the input, and we refer the reader to the comprehensive treatment of Kretschmann and Werner [26] for a discussion of all the concomitant ways of defining the capacity and the proof that asymptotically and for vanishing error they are the same. In the present paper we will measure the degree of approximation between states by the fidelity, given as

$$F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1 = \max |\langle \varphi | \psi \rangle|,$$

where the maximization is over all purifications  $|\varphi\rangle, |\psi\rangle$  of  $\rho$  and  $\sigma$ , respectively [24], [47]. This definition extends to subnormalized density operators  $\rho, \sigma \in \mathcal{S}_{\leq}(\mathcal{H})$  by letting

$$\begin{aligned}F(\rho, \sigma) &:= F(\rho \oplus (1 - \text{Tr } \rho), \sigma \oplus (1 - \text{Tr } \sigma)) \\ &= \|\sqrt{\rho}\sqrt{\sigma}\|_1 + \sqrt{(1 - \text{Tr } \rho)(1 - \text{Tr } \sigma)}.\end{aligned}$$

It can be shown that both

$$\begin{aligned}P(\rho, \sigma) &:= \sqrt{1 - F(\rho, \sigma)^2}, \text{ and} \\ A(\rho, \sigma) &:= \arccos F(\rho, \sigma) = \arcsin P(\rho, \sigma),\end{aligned}$$

called the *purified distance* and the *geodesic distance*, respectively, are metrics on  $\mathcal{S}_{\leq}(\mathcal{H})$ , cf. [45]. They are obviously equivalent, and can be shown to be equivalent to the trace norm distance [22]:

$$\frac{1}{2}\|\rho - \sigma\|_1 \leq P(\rho, \sigma) \leq \sqrt{\|\rho - \sigma\|_1}. \quad (1)$$

In the subsequent definitions, we will consistently use the purified distance. For instance, the error of a code  $(\mathcal{E}, \mathcal{D})$  for  $\mathcal{N}$  is defined as

$$P(\text{id}, \mathcal{D} \circ \mathcal{N} \circ \mathcal{E}) := \sup_{C'} \sup_{\rho \in \mathcal{S}(CC')} P(\rho, (\text{id} \otimes \mathcal{D} \circ \mathcal{N} \circ \mathcal{E})\rho).$$

The maximum dimension  $|C|$  of  $C$  such that there exists a quantum code for  $\mathcal{N}^{\otimes n}$  with error  $\epsilon$ , is denoted  $N(n, \epsilon)$ , or more precisely  $N(n, \epsilon|\mathcal{N})$  if we want to refer explicitly to the channel.

If we have a code with error  $\leq \epsilon$ , this means that we can use it with the maximally entangled state  $|\Phi\rangle^{CC'}$  at the input, to get an output state

$$\sigma^{CC'} = (\text{id} \otimes \mathcal{D} \circ \mathcal{N} \circ \mathcal{E})\Phi = (\text{id} \otimes \mathcal{D} \circ \mathcal{N})(\text{id} \otimes \mathcal{E})\Phi,$$

which is  $\epsilon$ -close to being maximally entangled:  $P(\Phi, \sigma) \leq \epsilon$ . This motivates the definition of an *entanglement-generating code with error  $\epsilon$* , which consists of a state  $\rho^{A'C'}$  and a decoding cptp map  $\mathcal{D} : \mathcal{L}(B) \rightarrow \mathcal{L}(C)$ , such that

$$P(\Phi^{CC'}, (\text{id} \otimes \mathcal{D} \circ \mathcal{N})\rho^{A'C'}) \leq \epsilon.$$

The maximum dimension  $|C|$  of  $C$  such that there exists an entanglement-generating code for  $\mathcal{N}^{\otimes n}$  with error  $\epsilon$ , is denoted  $N_E(n, \epsilon)$ , or more explicitly,  $N_E(n, \epsilon|\mathcal{N})$ . Clearly,  $N(n, \epsilon) \leq N_E(n, \epsilon)$ .

**Remark** Since the purified distance  $P(\Phi, (\text{id} \otimes \mathcal{D} \circ \mathcal{N})\rho) = \sqrt{1 - \text{Tr}((\text{id} \otimes \mathcal{D} \circ \mathcal{N})\rho)\Phi}$  is concave in  $\rho$ , we may always assume that the state  $\rho$  on  $A'C'$  in an entanglement-generating code is pure, as in each convex decomposition of  $\rho$  there is at least one state with an error no larger than that of  $\rho$ . ■

The quantum capacity is now defined as

$$Q(\mathcal{N}) = \inf_{\epsilon > 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log N(n, \epsilon).$$

One obtains the same capacity when using  $\limsup$  and  $N_E$ , see [26] for a proof of this and the equivalence of other variations of the definition.

A Shannon-style formula for the quantum capacity was first stated by Lloyd [27] and proved rigorously by Shor [38] and Devetak [15]. More precisely, in these papers they prove the direct (achievability) part which together with the earlier result of Schumacher and Nielsen [34], [35], who showed the same quantity to be an upper bound (i.e., weak converse), leads to a formula for the quantum capacity. We expand upon this weak converse in the following section.

The formula for the quantum capacity is given in terms of the *coherent information*

$$I(A|B)_\rho = -S(A|B)_\rho = S(\rho^B) - S(\rho^{AB}),$$

where  $S(\rho) = -\text{Tr } \rho \log \rho$  is the von Neumann entropy, of a state  $\rho^{AB} = (\text{id} \otimes \mathcal{N})\phi^{AA'}$  with a “test state”  $\phi$  on  $AA'$ . Namely,

$$Q(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{1}{n} Q^{(1)}(\mathcal{N}^{\otimes n}),$$

with the single-letter expression

$$Q^{(1)}(\mathcal{N}) = \max_{\phi \in \mathcal{S}(AA')} \{I(A|B)_\rho : \rho = (\text{id} \otimes \mathcal{N})\phi\}.$$

**Remark** The quantum capacity is known to be non-additive [44]. So is the single-letter quantity  $Q^{(1)}(\mathcal{N})$  [17], [40], meaning that the regularization above is necessary, at least as long as we base our capacity formula on the coherent information. It is not known whether there is a single-letter formula for  $Q(\mathcal{N})$ , or even an efficient approximation scheme [39]. As a matter of fact, we do not even know how to characterize the quantum capacity of the qubit depolarizing channel as a function of the noise. ■

### III. WEAK AND STRONG CONVERSE

The fact that the coherent information gives an upper bound on the quantum capacity of general channels has been known since Schumacher and Nielsen [34]. They showed that for any entanglement generating code with code space  $C$ , for a channel  $\mathcal{N} : \mathcal{L}(A') \rightarrow \mathcal{L}(B)$  with error  $\epsilon$ , using strong subadditivity together with Eq. (1) and the Fannes inequality, there exists an input test state  $\phi^{AA'}$  such that with  $\rho^{AB} = (\text{id} \otimes \mathcal{N})\phi$ ,

$$(1 - 2\epsilon) \log |C| \leq I(A)B)_\rho + 1.$$

Applying this to a maximal code for  $\mathcal{N}^{\otimes n}$  yields, for  $\epsilon < \frac{1}{2}$ ,

$$\frac{1}{n} \log N_E(n, \epsilon) \leq \frac{1}{1 - 2\epsilon} \frac{1}{n} Q^{(1)}(\mathcal{N}^{\otimes n}) + \frac{1}{(1 - 2\epsilon)n}, \quad (2)$$

hence the result that for  $n \rightarrow \infty$  and  $\epsilon \rightarrow 0$ , the optimal rate cannot exceed  $\lim_n \frac{1}{n} Q^{(1)}(\mathcal{N}^{\otimes n})$ , which we know is also asymptotically achievable, thanks to Lloyd-Shor-Devetak.

However, for any non-zero  $\epsilon > 0$ , the upper bound in Eq. (2) is a constant factor away from the capacity, which is the hallmark of a weak converse; it leaves room for a trade-off between communication rate and error, asymptotically.

If the quantum capacity  $Q(\mathcal{N})$  is zero, Eq. (2) says something a bit stronger, namely that  $N_E(n, \epsilon) \leq O(1)$ , at least when  $\epsilon < \frac{1}{2}$ . In this article we call such a statement *pretty strong converse*, i.e. a proof amounting to

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log N_E(n, \epsilon) \leq Q(\mathcal{N}),$$

at least for error  $\epsilon$  below some threshold  $\epsilon_0$ . A strong converse would require the above for all  $\epsilon < 1$ .

Here are two simple examples of channels for which the strong converse holds.

**Example (PPT entanglement binding channels).** If  $\mathcal{N}$  is such that all  $\rho = (\text{id} \otimes \mathcal{N})\phi$  have positive partial transpose (PPT), then any entanglement generating code for a maximally entangled state of Schmidt rank  $d$ , denoted  $\Phi_d$ , using any number  $n$  of channel uses and even arbitrary classical communication on the side, is still a PPT state. Twirling by the symmetries  $U \otimes \bar{U}$  of the maximally entangled state does not change the fidelity between the resulting state and the maximally entangled state. But the resulting isotropic state

$$\rho = p\Phi_d + (1 - p) \frac{1}{d^2 - 1} (\mathbb{1} - \Phi_d)$$

is still PPT, and it is well-known that this can only hold for  $p \leq \frac{1}{d}$  [30]. I.e., the error is at least  $\sqrt{1 - \frac{1}{d}}$ , which in the setting of  $n$  channel uses ( $\mathcal{N}^{\otimes n}$ ) goes to 1 exponentially fast for positive rates (meaning  $d = 2^{nR}$  for  $R > 0$ ). ■

**Example (Ideal channel).** Consider the identity  $\text{id}_2 : \mathcal{L}(\mathbb{C}^2) \rightarrow \mathcal{L}(\mathbb{C}^2)$  on a qubit and an entanglement-generating code for  $n$  uses of it,  $\text{id}_2^{\otimes n}$  for a maximally entangled state of rank  $d$ . It is evident that the state shared between sender and receiver after the transmission is of Schmidt rank  $\leq 2^n$ , and

so is any state obtained by the receiver's decoding. Hence the fidelity of the code is upper bounded by

$$\max \{ |\langle \Phi_d | \psi \rangle| : \text{Schmidt rank of } |\psi\rangle \text{ at most } 2^n \} = \sqrt{\frac{2^n}{d}}.$$

Consequently, as soon as the rate is above the capacity  $Q(\text{id}_2) = 1$ , i.e.  $d = 2^{nR}$  for  $R > 1$ , the error goes to 1 exponentially fast. ■

### IV. DEGRADABLE AND ANTI-DEGRADABLE CHANNELS

By the Stinespring dilation theorem, any channel can be defined by an isometric embedding  $U : A' \rightarrow B \otimes E$  followed by a partial trace over the environment system  $E$ , such that  $\mathcal{N}(\rho) = \text{Tr}_E U \rho U^\dagger$ . Tracing over  $B$  rather than  $E$  we obtain the corresponding complementary channel,  $\mathcal{N}^c(\rho) := \text{Tr}_B U \rho U^\dagger$ .

As we are interested in the channel's behaviour, we will without loss of generality assume from now on that  $E$  is chosen to be of minimal dimension (which makes  $U$  unique up to isometries on  $E$ ). Furthermore, since  $\mathcal{N}$  is the complementary channel of  $\mathcal{N}^c$ , we may equally reduce the dimension of  $B$  if necessary; this can equivalently be described as finding the subspace  $\hat{B} \subset B$  that contains all supports of all  $\mathcal{N}(\rho)$  for states  $\rho$  on  $A'$ , which is in fact the supporting subspace of  $\mathcal{N}(\mathbb{1})$ , and viewing  $\mathcal{N}$  as a mapping into  $\mathcal{L}(\hat{B})$ .

A channel  $\mathcal{N}$  is called *degradable* if it can be degraded to its complementary channel, i.e. if there exists a cptp map  $\mathcal{M}$  such that  $\mathcal{N}^c = \mathcal{M} \circ \mathcal{N}$ . Introducing the Stinespring dilation of  $\mathcal{M}$  by an isometry  $V : B \rightarrow F \otimes E'$ , the channel output system  $B$  can be mapped to the composite system  $E' \otimes F$  such that the channel taking  $A'$  to  $E$  is the same as the channel taking  $A'$  to  $E'$  (with an isomorphism between  $E$  and  $E'$  fixed once and for all). We may also assume  $F$  to be minimal. The above information process is illustrated in Fig. 1.

If the complementary channel is degradable, i.e. if  $\mathcal{N} = \mathcal{M} \circ \mathcal{N}^c$  for some cptp map, we call  $\mathcal{N}$  *anti-degradable*. A channel that is both degradable and anti-degradable is called *symmetric* [41].

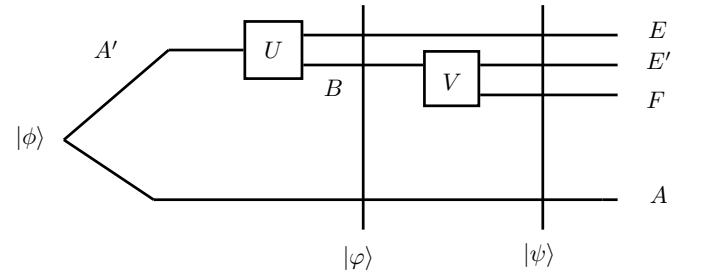


Fig. 1. Schematic of a degradable quantum channel, with the input state  $\phi$  between  $A'$  and the reference  $A$ , the channel output and environment state  $\varphi$  and the state  $\psi$  shared between  $A$ ,  $F$  and the two copies of the original environment,  $E$  and  $E'$ .

The identity between the channels  $\mathcal{L}(A') \rightarrow \mathcal{L}(E)$  and  $\mathcal{L}(A') \rightarrow \mathcal{L}(E')$  (defined by conjugating by  $VU$  and tracing over  $E'F$  and  $EF$ , respectively) is expressed by the equation

$$\psi^{AE} = \psi^{AE'}, \quad (3)$$

modulo the implicit isomorphism between  $E$  and  $E'$ . This was enough for Devetak and Shor [16] to prove that for degradable channels the coherent information is additive; see also [11, Sec. A.2]. The crucial point in their argument is that the coherent information can be rewritten as a conditional entropy,

$$I(A|B)_\varphi = S(F|E')_\psi. \quad (4)$$

Then, based on the observation that the state  $\psi^{FE'}$  on the r.h.s. is a linear function of the reference state  $\rho^{A'} = \text{Tr}_{A\phi}$ , and using strong subadditivity, one gets subadditivity of the coherent information of a product channel, hence additivity of  $Q^{(1)}$ . Below we give an alternative account of the reasoning leading to Eq. (4), which while being more complicated than those cited, has the benefit of suggesting an extension to min-entropies (Section V). For the class of degradable channels it is also known that the quantum capacity equals the private capacity [43] – see Section VI below.

Denoting  $\text{SWAP}_{EE'}$  the swap unitary between systems  $E$  and  $E'$ , i.e.  $\text{SWAP}_{EE'}|u\rangle|v\rangle = |v\rangle|u\rangle$  (always modulo the implicit identification of  $E$  with  $E'$ ), we have the following statement strengthening Eq. (3):

**Lemma 1** *Consider a degradable channel  $\mathcal{N}$  with Stinespring dilation  $U : A' \hookrightarrow B \otimes E$ . Then there exists a degrading map  $\mathcal{M}$  with Stinespring dilation  $V : B \hookrightarrow F \otimes E'$  (not necessarily with minimal dimension  $|F|$ ) and a unitary  $X$  on  $F$ , which may be chosen as an involution (i.e.  $X^2 = \mathbb{I}$ ), such that*

$$(X_F \otimes \text{SWAP}_{EE'})VU = VU.$$

*In particular, for arbitrary state vector  $|\phi\rangle^{AA'}$  and  $|\psi\rangle^{AFEE'} := (\mathbb{I} \otimes VU)|\phi\rangle$ ,*

$$(\mathbb{I}_A \otimes X_F \otimes \text{SWAP}_{EE'})|\psi\rangle^{AFEE'} = |\psi\rangle^{AFEE'}.$$

*Proof:* Start with an arbitrary dilation  $V_0 : B \hookrightarrow F_0 \otimes E'$  of an arbitrary map  $\mathcal{M}_0$ , and define the following isometry  $W : A \hookrightarrow EE'FG$ ,

$$W := \frac{1}{\sqrt{2}}(V_0U \otimes |0\rangle^G + \text{SWAP}_{EE'}V_0U \otimes |1\rangle^G),$$

with a qubit system  $G$ . Let  $F = F_0 \otimes G$  and  $X_F := \mathbb{I}_{F_0} \otimes X_G$ , where  $X$  is the Pauli  $\sigma_x$  unitary on  $G$ . Evidently,

$$W = (\text{SWAP}_{EE'} \otimes X_F)W,$$

and also, since  $\mathcal{N}$  is degradable,

$$\text{Tr}_{E'F} W \rho W^\dagger = \mathcal{N}^c(\rho).$$

Hence, the Stinespring dilations  $U$  and  $W$  are equivalent; to be precise, there exists an isometry  $V : B \hookrightarrow E'F$  such that  $W = VU$ , and we get  $VU = (\text{SWAP}_{EE'} \otimes X_F)VU$ . ■

The following reasoning uses the chain rule identity  $S(AB|C) = S(B|C) + S(A|BC)$  of the conditional von Neumann entropy, but no explicit expansion of any conditional entropy as a difference of two entropies. Consider a generic input state  $\phi^{AA'}$  to  $\mathcal{N}$  and its associated  $\varphi^{ABE}$  and  $\psi^{AFEE'}$ . Now, by invariance of the conditional entropy  $S(A|B) = S(AB) - S(B)$  under local unitaries and the duality identity

$S(A|B) = -S(A|C)$  with respect to a pure state on  $ABC$ , combined with the above lemma,

$$\begin{aligned} I(A|B)_\varphi &= -S(A|B)_\varphi \\ &= -S(A|FE')_\psi \\ &= S(F|E') - S(AF|E') \\ &= S(F|E') + S(AF|E) \\ &= S(F|E') + S(AF|E'). \end{aligned}$$

This shows that  $S(AF|E) = 0$ , and we obtain Eq. (4).

## V. PRETTY STRONG CONVERSE

**Theorem 2** *Let  $\mathcal{N} : \mathcal{L}(A) \rightarrow \mathcal{L}(B)$  be a degradable channel with finite quantum systems  $A$  and  $B$ . Then, there exists a constant  $\mu$  such that for error  $\epsilon < \frac{1}{\sqrt{2}}$  and every integer  $n$ ,*

$$\begin{aligned} \log N(n, \epsilon) &\leq \log N_E(n, \epsilon) \\ &\leq nQ^{(1)}(\mathcal{N}) + \mu \sqrt{n \ln \frac{64n|A|^2}{\lambda^2}} \\ &\quad + 3|A|^2 \log n + 5 + 5 \log \frac{1}{\lambda}, \\ &\leq nQ^{(1)}(\mathcal{N}) + O\left(\sqrt{n \log n}\right), \end{aligned}$$

where  $\lambda = \frac{1}{4} \left( \frac{1}{\sqrt{2}} - \epsilon \right)$ .

Together with the direct part (achievability proved in [15], [27], [38]) we thus get:

**Corollary 3** *For a degradable channel  $\mathcal{N}$ , the quantum capacity is given by*

$$\begin{aligned} Q(\mathcal{N}) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log N(n, \epsilon) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log N_E(n, \epsilon), \end{aligned}$$

for any  $0 < \epsilon < \frac{1}{\sqrt{2}}$ . Compared to the original definition this is simpler as we do not need to vary  $\epsilon$ , and there is convergence rather than reference to  $\liminf$  or  $\limsup$ . ■

The proof of this theorem will rely on the calculus of min- and max-entropies, of which we will briefly review the necessary definitions and properties; we refer the reader to [45] for more details.

**Definition 4 (Min- and max-entropy)** *For  $\rho^{AB} \in \mathcal{S}_{\leq}(AB)$ , the min-entropy of  $A$  conditioned on  $B$  is defined as*

$$H_{\min}(A|B)_\rho := \max_{\sigma_B \in \mathcal{S}(B)} \max\{\lambda \in \mathbb{R} : \rho^{AB} \leq 2^{-\lambda} \mathbb{I} \otimes \sigma^B\}.$$

*With a purification  $|\psi\rangle^{ABC}$  of  $\rho$ , we define*

$$H_{\max}(A|B)_\rho := -H_{\min}(A|C)_{\psi^{AC}},$$

*with the reduced state  $\psi^{AC} = \text{Tr}_B \psi$ .*

**Definition 5 (Smooth min- and max-entropy)** *Let  $\epsilon \geq 0$  and  $\rho_{AB} \in \mathcal{S}(AB)$ . The  $\epsilon$ -smooth min-entropy of  $A$  conditioned on  $B$  is defined as*

$$H_{\min}^\epsilon(A|B)_\rho := \max_{\rho' \approx_\epsilon \rho} H_{\min}(A|B)_{\rho'},$$

where  $\rho' \approx_\epsilon \rho$  means  $P(\rho', \rho) \leq \epsilon$  for  $\rho' \in \mathcal{S}_\leq(AB)$ .

Similarly,

$$\begin{aligned} H_{\max}^\epsilon(A|B)_\psi &:= \min_{\rho' \approx_\epsilon \rho} H_{\max}(A|B)_{\rho'} \\ &= -H_{\min}^\epsilon(A|C)_\psi, \end{aligned}$$

with a purification  $\psi \in \mathcal{S}(ABC)$  of  $\rho$ .

All min- and max-entropies, smoothed or not, are invariant under local unitaries and local isometries.

**Lemma 6 (Monotonicity)** For a state  $\rho \in \mathcal{S}(ABC)$  and any  $\epsilon \geq 0$ ,

$$\begin{aligned} H_{\min}^\epsilon(A|BC) &\leq H_{\min}^\epsilon(A|B), \\ H_{\max}^\epsilon(A|BC) &\leq H_{\max}^\epsilon(A|B). \end{aligned}$$

Since every cptp map can be written as an isometry followed by a partial trace, this means that for every  $\rho \in \mathcal{S}(AB)$  and cptp map  $\mathcal{T} : \mathcal{L}(B) \rightarrow \mathcal{L}(C)$ ,

$$\begin{aligned} H_{\min}^\epsilon(A|B)_\rho &\leq H_{\min}^\epsilon(A|C)_{(\text{id} \otimes \mathcal{T})\rho}, \\ H_{\max}^\epsilon(A|B)_\rho &\leq H_{\max}^\epsilon(A|C)_{(\text{id} \otimes \mathcal{T})\rho}. \end{aligned}$$

■

The following relations generalize the well-known chain rule identity  $S(AB|C) = S(B|C) + S(A|BC)$  for the von Neumann entropy, albeit for min- and max-entropies it turns into one of a set of inequalities. There are eight versions of it [48], of which we cite only the two we are going to use.

**Lemma 7 (Chain rules [21], [48])** Let  $\epsilon, \delta \geq 0$ ,  $\eta > 0$ . Then, with respect to the same state  $\rho \in \mathcal{S}(ABC)$ ,

$$\begin{aligned} H_{\max}^{\epsilon+2\delta+\eta}(AB|C) &\leq H_{\max}^\delta(B|C) + H_{\max}^\epsilon(A|BC) \\ &\quad + \log \frac{2}{\eta^2}, \end{aligned} \quad (5)$$

and

$$\begin{aligned} H_{\max}^\epsilon(AB|C) &\geq H_{\min}^\delta(B|C) + H_{\max}^{\epsilon+2\delta+2\eta}(A|BC) \\ &\quad - 3 \log \frac{2}{\eta^2}. \end{aligned} \quad (6)$$

■

**Lemma 8 (Proposition 5.5 in [45])** Let  $\rho \in \mathcal{S}(AB)$  and  $\alpha, \beta \geq 0$  such that  $\alpha + \beta < \frac{\pi}{2}$ . Then,

$$H_{\min}^{\sin \alpha}(A|B)_\rho \leq H_{\max}^{\sin \beta}(A|B)_\rho + \log \frac{1}{\cos^2(\alpha + \beta)}. \quad (7)$$

For  $\epsilon, \delta \geq 0$ ,  $\epsilon + \delta < 1$  this can be relaxed to the simpler form

$$H_{\min}^\epsilon(A|B)_\rho \leq H_{\max}^\delta(A|B)_\rho + \log \frac{1}{1 - (\epsilon + \delta)^2}. \quad (8)$$

■

**Lemma 9 (Dupuis [20])** Let  $\rho \in \mathcal{S}(AB)$  and  $0 \leq \epsilon \leq 1$ . Then,

$$H_{\max}^{\sqrt{1-\epsilon^4}}(A|B)_\rho \leq H_{\min}^\epsilon(A|B)_\rho, \quad (9)$$

which can be rewritten and relaxed into the form

$$\begin{aligned} H_{\max}^\delta(A|B)_\rho &\leq H_{\min}^{\sqrt[4]{1-\delta^2}}(A|B)_\rho \\ &\leq H_{\min}^{1-\frac{1}{4}\delta^2}(A|B)_\rho, \end{aligned} \quad (10)$$

for  $0 \leq \delta \leq 1$ . ■

*Proof of Theorem 2:* Consider an entanglement generation code for  $\log N_E(n, \epsilon)$  ebits of error  $\epsilon$  for the channel  $\mathcal{N}^{\otimes n}$ . As observed in conjunction with the definitions,  $N(n, \epsilon) \leq N_E(n, \epsilon)$  and w.l.o.g. the input state  $\phi^{\tilde{A}:A'^n}$  to the entanglement-generating code is pure (see Remark in Section II). Similar to Fig. 1, write

$$\begin{aligned} |\varphi\rangle^{\tilde{A}B^n E^n} &:= (\mathbb{1} \otimes U^{\otimes n})|\phi\rangle, \\ |\psi\rangle^{\tilde{A}E'^n F^n E^n} &:= (\mathbb{1} \otimes (V \otimes \mathbb{1})^{\otimes n})|\varphi\rangle. \end{aligned}$$

By definition, there exists a decoding cptp map  $\mathcal{D} : \mathcal{L}(B^n) \rightarrow \mathcal{L}(\tilde{A}')$ , such that  $\sigma = (\text{id} \otimes \mathcal{D} \circ \mathcal{N})\phi$  has purified distance  $\leq \epsilon$  from the maximally entangled state  $\Phi_{\tilde{A}\tilde{A}'}$ . Note that  $|\tilde{A}| = |\tilde{A}'| = N_E(n, \epsilon)$ . Hence, by definition of the max-entropy and using its monotonicity under cptp maps (Lemma 6),

$$\begin{aligned} \log N_E(n, \epsilon) &\leq -H_{\max}^\epsilon(\tilde{A}|\tilde{A}')_\sigma \\ &\leq -H_{\max}^\epsilon(\tilde{A}|B^n)_\varphi \\ &= -H_{\max}^\epsilon(\tilde{A}|E'^n F^n)_\psi. \end{aligned}$$

The latter, by the duality relation (Definition 5), is equal to  $H_{\min}^\epsilon(\tilde{A}|E)$ , which relates the coding performance directly to the decoupling principle (cf. [19]). But we shall not use that route and instead invoke the chain rule [Lemma 7, Eq. (5)], with  $\eta = \lambda = \frac{1}{4} \left( \frac{1}{\sqrt{2}} - \epsilon \right)$ , to continue

$$\begin{aligned} \log N_E(n, \epsilon) &\leq H_{\max}^\lambda(F^n|E'^n) \\ &\quad - H_{\max}^{\epsilon+3\lambda}(\tilde{A}F^n|E'^n) + \log \frac{2}{\lambda^2}. \end{aligned} \quad (11)$$

Let us deal with the second term here first: Using duality, and invoking Lemma 8, Eq. (7) with  $\alpha = \beta = \arcsin(\epsilon + 3\lambda) < \frac{\pi}{4}$ , we get

$$\begin{aligned} -H_{\max}^{\epsilon+3\lambda}(\tilde{A}F^n|E'^n) &= H_{\min}^{\sin \alpha}(\tilde{A}F^n|E^n) \\ &\leq H_{\max}^{\sin \alpha}(\tilde{A}F^n|E^n) + \log \frac{1}{\cos^2(2\alpha)} \\ &= H_{\max}^{\sin \alpha}(\tilde{A}F^n|E'^n) + 2 \log \frac{1}{\cos(2\alpha)}, \end{aligned}$$

using the symmetry of  $\psi$  with respect to swapping  $E^n$  and  $E'^n$ , as expressed in Lemma 1. We find that

$$\begin{aligned} -H_{\max}^{\epsilon+3\lambda}(\tilde{A}F^n|E'^n) &\leq \log \frac{1}{1 - 2(\epsilon + 3\lambda)^2} \\ &= \log \frac{1}{1 - 2\left(\frac{1}{\sqrt{2}} - \lambda\right)^2} \leq \log \frac{1}{2\lambda}. \end{aligned} \quad (12)$$

Turning to the first term in Eq. (11), we note that it is evaluated on  $\psi^{F^n E'^n} = V^{\otimes n} \mathcal{N}^{\otimes n}(\rho^{(n)}) V^{\dagger \otimes n}$ , a linear function of the input density  $\rho^{(n)} = \text{Tr}_{\tilde{A}} \phi \in \mathcal{S}(A'^n)$ . By slight abuse of notation we henceforth write

$$H_{\max}^\lambda(F^n|E'^n)_{\rho^{(n)}} = H_{\max}^\lambda(F^n|E'^n)_\psi.$$

Now, if we knew that the maximum of this max-entropy is attained on a tensor power state  $\rho^{(n)} = \rho^{\otimes n}$ , then we would be done, by immediately applying the asymptotic equipartition property (AEP) for min- and max-entropies (Proposition 13). A priori, however, the state  $\rho^{(n)}$  is arbitrary (note that it eventually comes directly from the optimal code with which we started our reasoning), so we need to work a little more. To this end we shall exploit the permutation covariance of the channel; for any permutation  $\pi \in S_n$ , acting naturally on an  $n$ -partite system, we have

$$\pi \psi^{(FE')^n} \pi^\dagger = V^{\otimes n} \mathcal{N}^{\otimes n} (\pi \rho^{(n)} \pi^\dagger) V^{\dagger \otimes n},$$

and since  $\pi^{(FE')^n} = \pi^{F^n} \otimes \pi^{E'^n}$  and by the local unitary invariance of the min- and max-entropies, we get

$$H_{\max}^\lambda(F^n|E'^n)_{\rho^{(n)}} = H_{\max}^\lambda(F^n|E'^n)_{\pi \rho^{(n)} \pi^\dagger}.$$

At this point we can use a restricted concavity property of the max-entropy, Lemma 10 below, and get

$$\begin{aligned} H_{\max}^\lambda(F^n|E'^n)_{\rho^{(n)}} &\leq H_{\max}^{\lambda/\sqrt{2}}(F^n|E'^n)_{\bar{\rho}^{(n)}} \\ &\leq H_{\min}^{1-\frac{1}{8}\lambda^2}(F^n|E'^n)_{\bar{\rho}^{(n)}}, \end{aligned} \quad (13)$$

for the permutation invariant state

$$\bar{\rho}^{(n)} = \frac{1}{n!} \sum_{\pi \in S_n} \pi \rho^{(n)} \pi^\dagger,$$

where we have also invoked Lemma 9, Eq. (10), in the second inequality in (13).

It is well-known that such permutation-invariant states are, in several meaningful senses, approximated by convex combinations of tensor power states; such a statement is known as (finite) de Finetti theorem, and here we use it in the form of the *Post-Selection Lemma* [10] (Lemma 12 below):<sup>1</sup>

$$\bar{\rho}^{(n)} \leq n^{|A|^2} \omega^{(n)},$$

where on the right we have the universal de Finetti state

$$\omega^{(n)} = \int d\sigma \sigma^{\otimes n},$$

for a certain universal measure on states  $\sigma \in \mathcal{S}(A)$ . Without loss of generality, by Carathéodory's Theorem, it may be assumed to be supported on  $M \leq n^{2|A|^2}$  points, hence we may write

$$\omega^{(n)} = \sum_{i=1}^M p_i \sigma_i^{\otimes n}.$$

Now we claim that

$$H_{\min}^{1-\frac{1}{8}\lambda^2}(F^n|E'^n)_{\bar{\rho}^{(n)}} \leq H_{\min}^{1-\frac{1}{16}\lambda^2 n^{-|A|^2}}(F^n|E'^n)_{\omega^{(n)}}. \quad (14)$$

Indeed, let  $\rho'$  be such that  $P(\rho', \bar{\rho}^{(n)}) \leq 1 - \delta := 1 - \frac{1}{8}\lambda^2$ . I.e., by the post-selection inequality and the operator monotonicity of the square root,

$$\begin{aligned} \sqrt{1 - (1 - \delta)^2} &\leq F(\rho', \bar{\rho}^{(n)}) = \left\| \sqrt{\rho'} \sqrt{\bar{\rho}^{(n)}} \right\|_1 \\ &\leq n^{\frac{1}{2}|A|^2} \left\| \sqrt{\rho'} \sqrt{\omega^{(n)}} \right\|_1, \end{aligned}$$

<sup>1</sup>We point out that it is also possible to do this using Renner's Exponential de Finetti Theorem [33], which requires a little more care to employ, but yields bounds quite similar to the ones obtained in the following.

thus

$$\begin{aligned} F(\rho', \omega^{(n)}) &\geq n^{-\frac{1}{2}|A|^2} \sqrt{2\delta - \delta^2} \geq \sqrt{\delta n^{-|A|^2}} \\ &\geq \sqrt{1 - (1 - \delta')^2}, \end{aligned}$$

with  $\delta' = \frac{1}{2}\delta n^{-|A|^2}$ . Hence, from Eqs. (13) and (14), Lemma 8, Eq. (8), and Lemma 11 below (with the finite-support decomposition of  $\omega^{(n)}$ ),

$$\begin{aligned} H_{\max}^\lambda(F^n|E'^n)_{\rho^{(n)}} &\leq H_{\max}^{\frac{1}{32}\lambda^2 n^{-|A|^2}}(F^n|E'^n)_{\omega^{(n)}} + \log \frac{32n^{|A|^2}}{\lambda^2} \\ &\leq \max_{\rho \in \mathcal{S}(A)} H_{\max}^{\frac{1}{32}\lambda^2 n^{-|A|^2}}(F^n|E'^n)_{\rho^{\otimes n}} \\ &\quad + 3|A|^2 \log n + 6 + \log \frac{1}{\lambda^2}. \end{aligned} \quad (15)$$

Putting Eqs. (11), (12) and (15) together, we arrive at

$$\begin{aligned} \log N_E(n, \epsilon) &\leq \max_{\rho \in \mathcal{S}(A)} H_{\max}^{\frac{1}{32}\lambda^2 n^{-|A|^2}}(F^n|E'^n)_{\rho^{\otimes n}} \\ &\quad + 3|A|^2 \log n + 5 + 5 \log \frac{1}{\lambda}. \end{aligned}$$

Note that the optimization over  $\rho$  is indeed a maximum since the smooth max-entropy is a continuous function of the state. The last step of the proof is an appeal to the quantum asymptotic equipartition property (Proposition 13),

$$H_{\max}^{\frac{1}{32}\lambda^2 n^{-|A|^2}}(F^n|E'^n)_{\rho^{\otimes n}} \leq nS(F|E')_\rho + \mu \sqrt{n \ln \frac{64n^{|A|^2}}{\lambda^2}},$$

and we are done. ■

**Remark** The error  $\frac{1}{\sqrt{2}}$  is precisely that achieved asymptotically by a single 50%-50% erasure channel acting on the code space, and of other suitable symmetric (i.e., degradable and anti-degradable) channels. We draw attention to the fact that in the proof we encounter a symmetric state, up to a local unitary,  $\psi^{\tilde{A}F^n:E^n:E'^n}$ , which can indeed be interpreted as the joint state between input ( $\tilde{A}F^n$ ), output ( $E^n$ ) and environment ( $E'^n$ ) of a suitable test state with a symmetric channel's Stinespring dilation.

We need to bound its min-entropy,  $H_{\min}^{\epsilon+3\lambda}(\tilde{A}F^n|E^n)$ , but if  $\epsilon \geq \frac{1}{\sqrt{2}}$ , then the overall smoothing parameter is strictly larger, and without any additional structure of the state we cannot upper bound the quantity further: Note that the symmetry we were using is consistent with an arbitrarily large entangled state passing through a single 50%-50% erasure channel. The smoothing by more than  $\frac{1}{\sqrt{2}}$  allows us to get rid of the erasure and pick out the successful transmission, yielding an arbitrarily large smooth min-entropy.

However, in Sections VII and VIII we will discuss other potential approaches, which might work because they use all the available structure. ■

Here are the lemmas needed in the above proof; they are proved in the appendix.

It is known that the max-entropy  $H_{\max}(A|B)_\rho$  is concave in the state  $\rho_{AB}$  [46], but this does not extend to the smoothed version. However, the following statement holds.

**Lemma 10** Let  $\rho \in \mathcal{S}(AB)$  be a state and consider the state family  $\rho_i^{AB} = (U_i \otimes V_i)\rho(U_i \otimes V_i)^\dagger$ , with unitaries  $U_i$  on  $A$  and  $V_i$  on  $B$ , and probabilities  $p_i$ ; define  $\bar{\rho} := \sum_i p_i \rho_i$ . Then,

$$H_{\max}^\epsilon(A|B)_{\bar{\rho}} \geq H_{\max}^{\epsilon\sqrt{2}}(A|B)_\rho.$$

**Lemma 11** For an ensemble  $\{p_i, \rho_i\}_{i=1}^M$  of states  $\rho_i \in \mathcal{S}(AB)$  with probabilities  $p_i$ , let  $\bar{\rho} = \sum_i p_i \rho_i$ . Then, for any  $0 \leq \epsilon \leq 1$ ,

$$H_{\max}^\epsilon(A|B)_{\bar{\rho}} \leq \max_i H_{\max}^\epsilon(A|B)_{\rho_i} + \log M.$$

**Lemma 12 (Post-Selection Technique [10])** For a Hilbert space  $\mathcal{H}$  of dimension  $d$ , denote by  $\text{Sym}^n(\mathcal{H})$  the subspace of permutation-invariant states in  $\mathcal{H}^{\otimes n}$ . Then, for every state  $\rho$  supported on  $\text{Sym}^n(\mathcal{H})$ ,

$$\rho \leq n^d \int d\psi |\psi\rangle\langle\psi|^{\otimes n} = P_{\text{Sym}^n(\mathcal{H})},$$

with the uniform (i.e., unitarily invariant) probability measure  $d\psi$  on pure states of  $\mathcal{H}$ , and – by Schur’s Lemma – the projector  $P_{\text{Sym}^n(\mathcal{H})}$  onto the symmetric subspace.

If  $\rho$  is a state on  $\mathcal{H}^{\otimes n}$  invariant under conjugation by permutations,  $\rho = \pi \rho \pi^\dagger$  for all  $\pi \in S_n$ , then the above can be applied to its purification in  $\text{Sym}^n(\mathcal{H} \otimes \mathcal{H}')$ , giving

$$\rho \leq n^{d^2} \int d\sigma \sigma^{\otimes n},$$

with a universal probability measure  $d\sigma$  on  $\mathcal{S}(\mathcal{H})$ . ■

Finally, we state a simplified version of the asymptotic equipartition property for min- and max-entropies, giving useful bounds for every  $n$ :

**Proposition 13 (Min- and max-entropy AEP [32], [45])**

Let  $\rho \in \mathcal{S}(\mathcal{H}_{AB})$  and  $0 < \epsilon < 1$ . Then,

$$\lim_{n \rightarrow \infty} \frac{1}{n} H_{\min}^\epsilon(A^n|B^n)_{\rho^{\otimes n}} = \lim_{n \rightarrow \infty} \frac{1}{n} H_{\max}^\epsilon(A^n|B^n)_{\rho^{\otimes n}} = S(A|B)_\rho.$$

More precisely, for a purification  $|\psi\rangle \in ABC$  of  $\rho$ , denote  $\mu_X := \log \|(\psi^X)^{-1}\|$ , where the inverse is the generalized inverse (restricted to the support), for  $X = B, C$ . Then, for every  $n$ ,

$$H_{\min}^\epsilon(A^n|B^n) \geq nS(A|B) - (\mu_B + \mu_C) \sqrt{n \ln \frac{2}{\epsilon}}, \quad (16)$$

$$H_{\max}^\epsilon(A^n|B^n) \leq nS(A|B) + (\mu_B + \mu_C) \sqrt{n \ln \frac{2}{\epsilon}}, \quad (17)$$

and similar opposite bounds via Lemma 8.

## VI. PRETTY STRONG CONVERSE FOR THE PRIVATE CAPACITY

In this section we show that the argument in the previous section can be augmented to yield a pretty strong converse for the private capacity.

We start by reviewing the basic definitions, which we adapt from Renes and Renner [31]: A *private classical code* for a channel  $\mathcal{N} : \mathcal{L}(A') \rightarrow \mathcal{L}(B)$  consists of a family of

signal states  $\rho_x \in \mathcal{S}(A')$  ( $x = 1, \dots, M$ ), and a decoding measurement (POVM)  $(D_x)_{x=1}^M$ , i.e.  $D_x \geq 0$ ,  $\sum_x D_x = \mathbb{1}_B$ . The latter can also be viewed as a cptp map  $\mathcal{D} : \mathcal{L}(B) \rightarrow \hat{\mathcal{X}}$ . Postulating a uniform distribution on the messages  $x$ , the code gives rise to the following averaged ccq-state of input, output and environment:

$$\sigma^{X\hat{X}E} = \frac{1}{M} \sum_x |x\rangle\langle x|^X \otimes |\hat{x}\rangle\langle \hat{x}|^{\hat{X}} \otimes \text{Tr}_B(V \rho_x V^\dagger)(D_{\hat{x}} \otimes \mathbb{1}_E),$$

encoding all correlations between legal users and eavesdropper of the system. The error of the code is defined in terms of the purified distance as

$$P\left(\frac{1}{M} \sum_x |x\rangle\langle x|^X \otimes |x\rangle\langle x|^{\hat{X}}, \sigma^{X\hat{X}}\right) = \sqrt{1 - \left(\frac{1}{M} \sum_x \sqrt{\text{Tr} \mathcal{N}(\rho_x) D_x}\right)^2}.$$

Its privacy is defined as

$$\min_{\tilde{\rho} \in \mathcal{S}(E)} P\left(\frac{1}{M} \sum_x |x\rangle\langle x|^X \otimes \tilde{\rho}^E, \sigma^{X\hat{X}}\right) = \min_{\tilde{\rho} \in \mathcal{S}(E)} \sqrt{1 - \left(\frac{1}{M} \sum_x F(\mathcal{N}^c(\rho_x), \tilde{\rho}^E)\right)^2}.$$

For a given channel  $\mathcal{N}$ , we denote the largest  $M$  such that there exists a private classical code with error  $\epsilon$  and privacy  $\delta$ , by  $M(n, \epsilon, \delta)$ . The (*weak*) *private capacity* of  $\mathcal{N}$  is then defined as

$$P(\mathcal{N}) = \inf_{\epsilon, \delta > 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log M(n, \epsilon, \delta).$$

It was determined in [9], [15], and like  $Q$  it is only known as a regularized characterization in general [42]. By the monogamy of entanglement, we know that  $P(\mathcal{N}) \geq Q(\mathcal{N})$  (see the Remark below), but in general this inequality is strict.

However for degradable channels, it was proved by Smith [43] that the private capacity  $P(\mathcal{N})$  equals the quantum capacity  $Q(\mathcal{N}) = Q^{(1)}(\mathcal{N})$ , and is hence given by a simple single-letter formula.

**Remark** The way we defined the code and the error above (as an average) is really that of a secret key generation code, analogous to the entanglement-generating codes in the previous section. To go from averaged error and privacy to essentially the same worst-case notions at the expense of loosing a constant fraction of the messages (hence no rate loss asymptotically) we use Ahlswede’s observation [2] on how randomization in the encoding can turn *several* average errors into only slightly worse worst-case errors.

For a code with messages  $x = 1, \dots, M$  and joint cq-state after decoding,

$$\rho^{ABE} = \frac{1}{M} \sum_{xy} P(y|x) |x\rangle\langle x| \otimes |y\rangle\langle y| \otimes \rho_{xy}^E,$$

consider the reduced states

$$\begin{aligned}\rho^{AB} &= \frac{1}{M} \sum_{xy} P(y|x) |x\rangle\langle x| \otimes |y\rangle\langle y|, \\ \rho^{AE} &= \frac{1}{M} \sum_x |x\rangle\langle x| \otimes \rho_x^E.\end{aligned}$$

With error and privacy are defined as above,

$$\epsilon = P\left(\rho^{AB}, \frac{1}{M} \sum_x |x\rangle\langle x| \otimes |x\rangle\langle x|\right)$$

and

$$\delta = P\left(\rho^{AE}, \frac{1}{M} \sum_x |x\rangle\langle x| \otimes \sigma^E\right),$$

where  $P = \sqrt{1 - F^2}$  is the purified distance, a short calculation shows that

$$\begin{aligned}F\left(\rho^{AB}, \frac{1}{M} \sum_x |x\rangle\langle x| \otimes |x\rangle\langle x|\right) &= \frac{1}{M} \sum_x \sqrt{P(x|x)} =: F_1, \\ F\left(\rho^{AE}, \frac{1}{M} \sum_x |x\rangle\langle x| \otimes \sigma^E\right) &= \frac{1}{M} \sum_x F(\rho_x^E, \sigma^E) =: F_2.\end{aligned}$$

We will now encode messages  $m$  into uniform distributions on pairwise disjoint sets  $K_m \subset [M] = \{1, \dots, M\}$  of cardinality  $k$ , with  $m = 1, \dots, N$  such that  $kN \leq M$ .

We will draw the elements of  $K_1, \dots, K_N$  randomly and without replacement from  $[M]$ . We then use Azuma's inequality to bound the probability that for a given  $m$  and  $\eta \geq 0$

$$\frac{1}{k} \sum_{x \in K_m} \sqrt{P(x|x)} < F_1 - \eta,$$

or

$$\frac{1}{k} \sum_{x \in K_m} F(\rho_x^E, \sigma^E) < F_2 - \eta.$$

Namely, each of these events has probability at most  $p = 2e^{-2k\eta^2}$  [3], [14]. The input-output-environment state of the new code for the messages  $m = 1, \dots, N$  is

$$\omega = \frac{1}{N} \sum_{mm'} \frac{1}{k} \sum_{x \in K_m, y \in K_{m'}} P(y|x) |m\rangle\langle m| \otimes |m'\rangle\langle m'| \otimes \rho_x^E.$$

Note that  $P(m|m) \geq \frac{1}{k} \sum_{x \in K_m} P(x|x)$ , and by concavity of the square root,

$$\sqrt{P(m|m)} \geq \frac{1}{k} \sum_{x \in K_m} \sqrt{P(x|x)}.$$

Likewise, the state of the eavesdropper for message  $m$  is  $\frac{1}{k} \sum_{x \in K_m} \rho_x^E$ , and by concavity of the fidelity,

$$F\left(\frac{1}{k} \sum_{x \in K_m} \rho_x^E, \sigma^E\right) \geq \frac{1}{k} \sum_{x \in K_m} F(\rho_x^E, \sigma^E).$$

I.e., this message will have error  $\leq \epsilon'$  and privacy deviation  $\leq \delta'$  for these “good”  $m$ , where it is straightforward to work out that  $\epsilon' \leq \epsilon(1 + \frac{\eta}{\epsilon^2})$  and  $\delta' \leq \delta(1 + \frac{\eta}{\delta^2})$ . In other words, by choosing  $\eta = a \cdot \min(\epsilon^2, \delta^2)$  we can make the new error and privacy arbitrarily close to the original parameters.

Now, we can find  $K_1, \dots, K_N$  such that a fraction  $\geq 1 - p$  of the  $K_m$  are “good”, throw away the “bad” ones and we are left with the code we want: it has  $N' \geq (1 - p)N = \frac{1-p}{k}M \geq \frac{1-p}{2k}M$  messages, if we choose  $k$  such that  $p \leq 1/2$ , which holds for  $k \geq \frac{\ln 4}{2\eta^2}$ .

In summary, we can get a code with randomized encoding and worst case error  $\epsilon' < (1 + a)\epsilon$ , worst case privacy  $\delta' < (1 + a)\delta$ , and losing a constant amount of information. Indeed the number of bits encoded diminishes by at most

$$2 \log \frac{1}{\eta} \leq 2 \log \frac{1}{a} + 4 \log \frac{1}{\epsilon} + 4 \log \frac{1}{\delta}.$$

■

By definition, every entanglement-generating code of error  $\epsilon'$  gives rise to a private classical (secret key generation) code of error and privacy  $\epsilon'$ , and with  $M = |C|$  messages. Thus,  $M(n, \epsilon', \epsilon') \geq N_E(n, \epsilon') \geq N(n, \epsilon')$ .

**Theorem 14** *Let  $\mathcal{N} : \mathcal{L}(A) \rightarrow \mathcal{L}(B)$  be a degradable channel with finite quantum systems  $A$  and  $B$ . Then, for error  $\epsilon$  and privacy  $\delta$  such that  $\epsilon + 2\delta < \frac{1}{\sqrt{2}}$  (e.g.  $\epsilon = \delta < \frac{1}{3\sqrt{2}} \approx .2357$ ), and every integer  $n$ ,*

$$\begin{aligned}\log M(n, \epsilon, \delta) &\leq nQ^{(1)}(\mathcal{N}) + \mu \sqrt{n \ln \frac{64n|A|^2}{\eta^2}} \\ &\quad + 3|A|^2 \log n + 9 + 11 \log \frac{1}{\eta}, \\ &\leq nQ^{(1)}(\mathcal{N}) + O\left(\sqrt{n \log n}\right),\end{aligned}$$

where  $\eta = \frac{1}{6} \left( \frac{1}{\sqrt{2}} - \epsilon - 2\delta \right)$ .

Together with the direct part (achievability proved in [9], [15]) we thus get:

**Corollary 15** *For a degradable channel  $\mathcal{N}$ , the private capacity is given by*

$$P(\mathcal{N}) = \lim_{n \rightarrow \infty} \frac{1}{n} \log M(n, \epsilon, \delta),$$

for any  $\epsilon, \delta > 0$  such that  $\epsilon + 2\delta < \frac{1}{\sqrt{2}}$ . ■

*Proof:* Consider a code for  $\mathcal{N}^{\otimes n}$  with  $M = M(n, \epsilon, \delta)$  messages, that has error  $\epsilon$  and is  $\delta$ -private: message  $x$  (chosen uniformly) is encoded as  $\sigma_x \in \mathcal{S}(A'^n)$  and sent through the channel, giving rise to an averaged cqg-state between reference  $X$ , output  $B^n$  and environment  $E^n$ :

$$\rho^{XB^nE^n} = \frac{1}{M} \sum_x |x\rangle\langle x|^X \otimes U^{\otimes n} \rho_x U^{\dagger \otimes n}.$$

The “trivial” converse shows that

$$\log M \leq H_{\min}^{\delta}(X|E^n) - H_{\max}^{\epsilon}(X|B^n),$$

cf. Renes and Renner [31], whose argument we briefly repeat here since they used trace norm rather than purified distance. According to the definition of privacy given above, the reduced state  $\rho^{XE^n}$  is within purified distance  $\delta$  of a product state of the form  $\frac{1}{M} \sum_x |x\rangle\langle x| \otimes \tilde{\rho}^{E^n}$ , hence  $H_{\min}^{\delta}(X|E^n) \geq \log M$ .



Likewise, there exists a decoding cptp map  $\mathcal{D} : \mathcal{L}(B^n) \rightarrow \hat{X}$  such that  $(\text{id} \otimes \mathcal{D})\rho^{XB^n}$  is within  $\epsilon$  purified distance from the perfectly correlated state  $\frac{1}{M} \sum_x |x\rangle\langle x|^X \otimes |x\rangle\langle x|^{\hat{X}}$ , hence  $H_{\max}^\epsilon(X|B^n) \leq 0$ .

Now we can purify  $\rho^{XB^n E^n} = \text{Tr}_{A_0 X'} \varphi^{X X' A_0 B^n E^n}$ , introducing a dummy system  $A_0$  to hold the purifications  $\phi_x^{A_0 A'^n}$  of the signal states  $\rho_x$  and a coherent copy  $X'$  of  $X$ :

$$|\varphi\rangle^{X X' B^n E^n} = \frac{1}{\sqrt{M}} \sum_x |x\rangle^X |x\rangle^{X'} (\mathbb{1}_{X X'} \otimes U^{\otimes n}) |\phi_x\rangle^{A_0 A'^n},$$

to which we then also apply the Stinespring dilation of the degrading map:

$$\begin{aligned} |\psi\rangle^{X X' E'^n F^n E^n} &= (\mathbb{1}_{X X'} \otimes V^{\otimes n} \otimes \mathbb{1}_{E^n}) |\varphi\rangle \\ &= \frac{1}{\sqrt{M}} \sum_x |x\rangle^X |x\rangle^{X'} (\mathbb{1}_{X X'} \otimes (VU)^{\otimes n}) |\phi_x\rangle^{A_0 A'^n}, \end{aligned}$$

With respect to  $\psi$ , we thus have

$$\begin{aligned} \log M &\leq H_{\min}^\delta(X|E^n) - H_{\max}^\epsilon(X|E'^n F^n) \\ &= H_{\min}^\delta(X|E'^n) - H_{\max}^\epsilon(X|E'^n F^n) \\ &\leq H_{\max}^\eta(F^n|E'^n) - H_{\max}^{\epsilon+2\delta+5\eta}(F^n|E'^n X) \quad (18) \\ &\quad + 4 \log \frac{2}{\eta^2}, \end{aligned}$$

where we have used the degradability property of the channel in the second line, and in the third line the chain rule, Lemma 7, in its two manifestations Eqs. (5) and (6). Indeed,

$$\begin{aligned} H_{\max}^{\epsilon+3\eta}(AB|C) &\leq H_{\max}^\eta(A|C) + H_{\max}^\epsilon(B|AC) + \log \frac{2}{\eta^2} \\ &\parallel \\ H_{\max}^\kappa(AB|C) &\geq H_{\min}^\delta(B|C) + H_{\max}^{\kappa+2\delta+2\eta}(A|BC) \\ &\quad - 3 \log \frac{2}{\eta^2}, \end{aligned}$$

which we employ with the identifications  $F^n \equiv A$ ,  $X \equiv B$ ,  $E'^n \equiv C$ , and with  $\kappa = \epsilon + 3\delta$ .

Choosing  $\eta = \frac{1}{6} \left( \frac{1}{\sqrt{2}} - \epsilon - 2\delta \right)$  ensures that  $\epsilon' := \epsilon + 2\delta + 5\eta = \frac{1}{\sqrt{2}} - \eta < \frac{1}{\sqrt{2}}$ , and we can bound the second term on the right hand side of Eq. (18) as before, in the proof of Theorem 2:

$$\begin{aligned} -H_{\max}^{\epsilon'}(F^n|E'^n X) &\leq -H_{\min}^{\epsilon'}(F^n|E'^n X) + 2 \log \frac{1}{2\eta} \\ &= H_{\max}^{\epsilon'}(F^n|E^n X' A_0) + 2 \log \frac{1}{2\eta} \\ &\leq H_{\max}^{\epsilon'}(F^n|E^n X') + 2 \log \frac{1}{2\eta} \\ &= H_{\max}^{\epsilon'}(F^n|E'^n X) + 2 \log \frac{1}{2\eta}, \end{aligned}$$

where we have used Lemma 8, then the duality between min- and max-entropy, then the monotonicity (Lemma 6) and finally the exchange symmetry between  $X$  and  $X'$  as well as between  $E$  and  $E'$ . As this means

$$-H_{\max}^{\epsilon'}(F^n|E'^n X) \leq \log \frac{1}{2\eta},$$

we have by plugging this into Eq. (18),

$$\log M(n, \epsilon, \delta) \leq H_{\max}^\eta(F^n|E'^n) + 3 + 9 \log \frac{1}{\eta},$$

and the rest of the argument is as in the proof of Theorem 2 [cf. Eq. (15)]:

$$\begin{aligned} H_{\max}^\eta(F^n|E'^n) &\leq H_{\max}^{\frac{1}{32}\eta^2 n^{-|A|^2}}(F^n|E'^n)_{\omega^{(n)}} + \log \frac{32n^{|A|^2}}{\eta^2} \\ &\leq \max_{\rho \in \mathcal{S}(A)} H_{\max}^{\frac{1}{32}\eta^2 n^{-|A|^2}}(F^n|E'^n)_{\rho^{\otimes n}} \\ &\quad + 3|A|^2 \log n + 6 + \log \frac{1}{\eta^2}, \end{aligned}$$

invoking the quantum AEP for the max-entropy (Proposition 13).  $\blacksquare$

## VII. STRONG CONVERSE FOR SYMMETRIC CHANNELS IMPLIES IT FOR DEGRADABLE CHANNELS

The main result of this section, Theorem 19, is valid for degradable channels satisfying the following technical condition.

**Definition 16** We say that a degradable channel  $\mathcal{N}$  is of type I (for invariance) if one can choose a Stinespring dilation  $U$  of it, and a Stinespring dilation  $V$  of a degrading channel  $\mathcal{M}$ , such that the unitary  $X_F$  in Lemma 1 is a global phase (hence  $\pm 1$ ). I.e.,

$$(\mathbb{1}_F \otimes \text{SWAP}_{EE'})UV = \pm UV.$$

**Example** (Erasure channels). The qubit erasure channel

$$\mathcal{E}_q(\rho) = (1-q)\rho \oplus q|*\rangle\langle*|$$

with erasure probability  $q \leq \frac{1}{2}$  has as its complementary channel  $\mathcal{E}_q^c = \mathcal{E}_{1-q}$ ; as degrading map serves  $\mathcal{E}_t$ , with  $t = \frac{q}{1-q}$  (augmented by the identity on  $|*\rangle\langle*|$ ).

We can guess an isometric dilation of  $\mathcal{E}_q$ ,

$$U : |\phi\rangle \mapsto \sqrt{1-q}|\phi\rangle^B |*\rangle^E + e^{i\alpha} \sqrt{q} |*\rangle^B |\phi\rangle^E,$$

and likewise for the degrading map,

$$\begin{aligned} V : |*\rangle &\mapsto |*\rangle^F |*\rangle^{E'} \\ |\phi\rangle &\mapsto \sqrt{1-t}|\phi\rangle^F |*\rangle^{E'} + \sqrt{t} |*\rangle^F |\phi\rangle^{E'}. \end{aligned}$$

With the choice of phase  $e^{i\alpha} = 1$ , it is straightforward to verify that  $\text{SWAP}_{EE'} VU = VU$ .

However, since the output of an erasure channel has no coherences between the erasure symbol and the unerased part, there is considerable freedom in choosing the dilations both of the channel and of the degrading map. For some of them there is no unitary  $X_F$  as in Lemma 1, for some the unitary is non-trivial. Indeed, we can see this by varying  $\alpha$  in the dilation  $U$  above, most choices of which leave no symmetry  $X_F$ , but for  $e^{i\alpha} = -1$  we can choose  $X_F = 2|*\rangle\langle*| - \mathbb{1}$ .  $\blacksquare$

**Example** (Schur multiplier channels). Given a positive semidefinite  $n \times n$ -matrix  $S \geq 0$  with diagonal entries  $S_{ii} = 1$  one can define a cptp map  $\mathcal{N}_S$  on  $n \times n$ -matrices by Schur/Hadamard multiplication of the input  $\rho$  by  $S$ :

$$\mathcal{N}_S : \rho \mapsto \rho \circ S, \text{ i.e. } \mathcal{N}_S(|i\rangle\langle j|) = S_{ij}|i\rangle\langle j|.$$

It is well-known that  $S$  can be viewed as Gram matrix of unit vectors  $|\varphi_1\rangle, \dots, |\varphi_n\rangle$ :

$$S_{ij} = \langle \varphi_j | \varphi_i \rangle,$$

suggesting a Stinespring dilation

$$U : |i\rangle \mapsto |i\rangle^B |\varphi_i\rangle^E.$$

It gives rise to the complementary channel

$$\mathcal{N}_S^c(|i\rangle\langle j|) = \delta_{ij} |\varphi_i\rangle\langle \varphi_i|,$$

so we can choose  $\mathcal{N}_S^c$  itself as degrading map and essentially  $U$  as its dilation  $V$  (with  $F$  taking the place of  $B$ , and  $E'$  that of  $E$ ).

Thus,

$$VU : |i\rangle \mapsto |i\rangle^F |\varphi_i\rangle^E |\varphi_i\rangle^{E'},$$

which is evidently invariant under  $\text{SWAP}_{EE'}$  since the output state restricted to  $EE'$ ,  $\text{Tr}_F VU \rho U^\dagger V^\dagger$ , is supported on the symmetric subspace of  $E \otimes E'$ . ■

**Remark** We do not know whether all degradable channels are of type I, not having found a counterexample so far. From the examples given above it is clear however that the dilations  $U$  and  $V$  required for a proof that a given channel is type I, have to be constructed carefully. The next lemma shows that for any degradable channel we can construct one that is information theoretically equivalent, and which is of type I. ■

**Lemma 17** *For every degradable channel  $\mathcal{N} : \mathcal{L}(A') \rightarrow \mathcal{L}(B)$ , the channel*

$$\begin{aligned} \tilde{\mathcal{N}} &= \mathcal{N} \otimes \tau^{B_0} : \mathcal{L}(A') \rightarrow \mathcal{L}(B \otimes B_0), \\ \rho &\mapsto \mathcal{N}(\rho) \otimes \tau^{B_0}, \end{aligned}$$

*which attaches to the output of  $\mathcal{N}$  a qubit system  $B_0$  in the maximally mixed state, is degradable of type I.*

*Proof:* Clearly,  $\tilde{\mathcal{N}}^c = \mathcal{N}^c \otimes \tau^{E_0}$ , with a qubit system  $E_0$ , so the new channel is also degradable.

Choose a Stinespring isometry  $U$  of  $\mathcal{N}$  and  $V$  of the degrading map  $\mathcal{M}$  according to Lemma 1, so that we have a unitary involution  $X_F$  with

$$(X_F \otimes \text{SWAP}_{EE'}) VU = VU.$$

$X_F$  can have only the two eigenvalues  $\pm 1$ , so decompose  $F = F_+ \oplus F_-$  into the respective eigenspaces with projectors  $P_+$  and  $P_-$ , respectively. Of course also  $\text{SWAP}_{EE'}$  has eigenvalues  $\pm 1$ , the corresponding eigenspaces being known as symmetric and anti-symmetric subspace, denoted as  $\text{Sym}^2(E)$  and  $\Lambda^2(E)$ , respectively.

The above invariance of  $VU$  under left multiplication by  $X_F \otimes \text{SWAP}_{EE'}$  is equivalently expressed by saying that  $VU$  maps  $A'$  into the  $+1$ -eigenspace of  $X_F \otimes \text{SWAP}_{EE'}$ , which is

$$F_+ \otimes \text{Sym}^2(E) \oplus F_- \otimes \Lambda^2(E).$$

In this picture we see why  $X_F$  is necessary: it is there to undo a possible phase of  $-1$  induced by  $\text{SWAP}_{EE'}$  (on  $\Lambda^2(E)$ ), by applying the same phase once more on  $F_-$ . We

can also see how to write down dilations of  $\tilde{\mathcal{N}}$  and a degrading map that avoid this problem: First,  $\tilde{U} : A' \hookrightarrow (B \otimes B_0) \otimes (E \otimes E_0)$  with

$$\tilde{U}|\phi\rangle := (U|\phi\rangle)^{BE} \otimes \left( \frac{|01\rangle + |10\rangle}{\sqrt{2}} \right)^{B_0 E_0}$$

is a dilation of  $\tilde{\mathcal{N}}$ . Secondly, we define a degrading map by writing down directly an isometric dilation  $\tilde{V} : B \otimes B_0 \hookrightarrow F \otimes (E' \otimes E'_0)$ :

$$\tilde{V}(|\varphi\rangle^B |b\rangle^{B_0}) := \left( \mathbb{1}_{E'} \otimes \text{C-Z}^{F \rightarrow E'_0} \right) ((V|\varphi\rangle)|b\rangle),$$

where

$$\text{C-Z}^{F \rightarrow E'_0} = P_+ \otimes \mathbb{1}_{B_0} + P_- \otimes Z_{B_0}$$

is a controlled-Z using the  $F_\pm$  subspaces to trigger a  $Z$  on the qubit  $E'_0$  (which we identify with  $B_0$ ).

It is easy to check that  $\text{Tr}_F \tilde{V} \cdot \tilde{V}^\dagger$  defines a bona fide degrading map for  $\tilde{\mathcal{N}}$ . But it is also of type I, as it can be confirmed by direct calculation that

$$\begin{aligned} \tilde{V} \tilde{U} |\phi\rangle &= (P_+ \otimes \mathbb{1}_{EE'}) VU |\phi\rangle \otimes \left( \frac{|01\rangle + |10\rangle}{\sqrt{2}} \right)^{E_0 E'_0} \\ &\quad + (P_- \otimes \mathbb{1}_{EE'}) VU |\phi\rangle \otimes \left( \frac{|01\rangle - |10\rangle}{\sqrt{2}} \right)^{E_0 E'_0}. \end{aligned}$$

Since the left hand factor in the first line is in  $F \otimes \text{Sym}^2(E)$ , while the analogous term in the second line is in  $F \otimes \Lambda^2(E)$ , the entire expression lies in  $F \otimes \text{Sym}^2(EE_0)$ , hence under the simultaneous swap  $EE_0 \leftrightarrow E'E'_0$ ,

$$\text{SWAP}_{EE_0: E'E'_0} \tilde{V} \tilde{U} = \tilde{V} \tilde{U},$$

and we are done. ■

Degradable channels of type I are intimately related to symmetric channels, as shown in the next lemma.

**Lemma 18** *Let  $\mathcal{N}$  be a degradable channel of type I, and choose a Stinespring dilation  $U$  as well as a dilation  $V$  of a degrading map, according to Lemma 1, s.t.  $X_F = \pm \mathbb{1}$ .*

*For any test state  $|\phi_0\rangle \in AA'$  of maximal Schmidt rank, let  $|\psi_0\rangle^{AFEE'} = VU|\phi_0\rangle^{AA'}$  and denote the supporting subspace of  $\psi_0^{AF}$  by  $G$ .*

*Then there is a symmetric channel  $\mathcal{M}$  with Stinespring isometry  $W : G' \hookrightarrow E \otimes E'$  (i.e.  $\text{SWAP}_{EE'} W = \pm W$ ) such that every state  $|\psi\rangle^{AFEE'} = VU|\phi\rangle$ , for  $|\phi\rangle \in AA'$  can be written as  $W|\xi\rangle \in GEE'$  for a suitable test state  $|\xi\rangle \in GG'$ , up to a (state-dependent) isometry  $\widehat{W} : G \hookrightarrow AF$ :*

$$|\psi\rangle^{AFEE'} = (\widehat{W} \otimes W) |\xi\rangle^{GG'}.$$

*Proof:* By definition,  $|\psi_0\rangle^{AFEE'} \in G \otimes E \otimes E'$ , so we may denote it as well  $|\psi_0\rangle^{GEE'}$ . Choose a purification  $|\chi\rangle^{GG'}$  with  $G' \simeq G$ , so that there exists an isometry  $W : G' \hookrightarrow EE'$  with

$$(\mathbb{1} \otimes W) |\chi\rangle^{GG'} = |\psi_0\rangle^{GEE'}.$$

It is easy to see that  $W$  has the required symmetry property: since  $\text{SWAP}_{EE'} |\psi_0\rangle^{GEE'} = \pm |\psi_0\rangle^{GEE'}$ , it follows that  $(\mathbb{1} \otimes$

$\text{SWAP}_{EE'}W)|\chi\rangle^{GG'} = \pm(\mathbb{1} \otimes W)|\chi\rangle^{GG'}$ , and since  $|\chi\rangle$  has maximal Schmidt rank,  $\text{SWAP}_{EE'}W = \pm W$  follows.

Now, let  $|\phi\rangle^{AA'}$  be an arbitrary input test state and  $|\psi\rangle^{AFEE'} = VU|\phi\rangle$ . Then,

$$|\phi\rangle^{AA'} = \left( \sqrt{\phi^A} \sqrt{\phi_0^A}^{-1} \otimes \mathbb{1} \right) |\phi_0\rangle^{AA'},$$

and thus

$$\begin{aligned} |\psi\rangle^{AFEE'} &= \left( \sqrt{\phi^A} \sqrt{\phi_0^A}^{-1} \otimes \mathbb{1}^F \otimes \mathbb{1}^{EE'} \right) |\psi_0\rangle^{AFEE'} \\ &= \left( \sqrt{\psi^{AF}} \sqrt{\psi_0^{AF}}^{-1} \otimes \mathbb{1}^{EE'} \right) |\psi_0\rangle^{AFEE'} \\ &= \left( \sqrt{\psi^{AF}} \sqrt{\psi_0^{AF}}^{-1} \otimes W \right) |\chi\rangle^{GG'}. \end{aligned}$$

Finally, since  $\chi^G = \psi_0^{AF}$  has support on  $G$ , there exists a  $\sigma \in \mathcal{S}(G)$  and an isometry  $\widehat{W} : G \hookrightarrow AF$  such that

$$\begin{aligned} &\left( \sqrt{\psi^{AF}} \sqrt{\phi_0^{AF}}^{-1} \otimes \mathbb{1}^{G'} \right) |\chi\rangle^{GG'} \\ &= \left( \widehat{W} \sqrt{\sigma} \sqrt{\chi_G}^{-1} \otimes \mathbb{1}^{G'} \right) |\chi\rangle^{GG'} \\ &=: (\widehat{W} \otimes \mathbb{1}^{G'}) |\xi\rangle^{GG'}. \end{aligned}$$

In total,  $|\psi\rangle^{AFEE'} = (\widehat{W} \otimes W) |\xi\rangle^{GG'}$ , which is what we wanted to prove.  $\blacksquare$

**Theorem 19** *Let  $\mathcal{N} : \mathcal{L}(A) \rightarrow \mathcal{L}(B)$  be a degradable channel, which w.l.o.g. we assume to be of type I (by Lemma 17). Denote its environment by  $E$  and the associated symmetric channel by  $\mathcal{M}$ , with Stinespring dilation  $W : G \hookrightarrow E \otimes E'$  from Lemma 18. Then  $\mathcal{N}$  obeys the strong converse for its quantum capacity, if  $\mathcal{M}$  does (note that by the no-cloning argument,  $Q(\mathcal{M}) = 0$ ). More precisely, there exists a constant  $\mu$  such that*

$$\begin{aligned} \log N_E(n, \epsilon | \mathcal{N}) &\leq nQ^{(1)}(\mathcal{N}) + \mu \sqrt{n \ln \frac{64n|A|^2}{\lambda^2}} \\ &\quad + 8 \log \frac{1}{\lambda} + O(\log n) \\ &\quad + \log N_E(n, 1 - \lambda | \mathcal{M}), \end{aligned}$$

with  $\lambda = \frac{1-\epsilon}{5}$ .

*Proof:* We follow the initial steps of the proof of Theorem 2, until the bound

$$\log N_E(n, \epsilon) \leq H_{\max}^\lambda(F^n | E'^n) - H_{\max}^{\epsilon+3\lambda}(AF^n | E'^n) + \log \frac{2}{\lambda^2},$$

where all entropies are with respect to the state  $|\psi\rangle^{AE^n F^n E'^n}$ . Now we choose  $\lambda = \frac{1-\epsilon}{5}$ .

The first term is treated in the exact same way as we did there, giving

$$\begin{aligned} H_{\max}^\lambda(F^n | E'^n) &\leq \max_{\rho \in \mathcal{S}(A)} H_{\max}^{\frac{1}{32}\lambda^2 n^{-|A|^2}}(F^n | E'^n)_{\rho^{\otimes n}} \\ &\quad + 3|A|^2 \log n + 6 + \log \frac{1}{\lambda^2} \\ &\leq nQ^{(1)}(\mathcal{N}) + \mu \sqrt{n \ln \frac{64n|A|^2}{\lambda^2}} \\ &\quad + 3|A|^2 \log n + 6 + \log \frac{1}{\lambda^2}, \end{aligned}$$

where we have used the quantum AEP (Proposition 13) once more.

The second term can be upper bounded

$$\begin{aligned} -H_{\max}^{\epsilon+3\lambda}(AF^n | E'^n) &= H_{\min}^{\epsilon+3\lambda}(AF^n | E'^n)_\psi \\ &= H_{\min}^{\epsilon+3\lambda}(G^n | E^n)_{(\mathbb{1} \otimes W)^{\otimes n} |\xi\rangle} \\ &\leq \log N_E(n, \epsilon + 4\lambda | \mathcal{M}) + 4 \log \frac{1}{\lambda}, \end{aligned}$$

using duality in the first equation and Lemma 18 in the second, to rewrite the state  $|\psi\rangle^{AF^n E^n E'^n}$  (up to an isometry  $G^n \hookrightarrow AF^n$ ) as if a test state  $|\xi\rangle^{G^n G'^n}$  had gone through  $W^{\otimes n}$ . The inequality in the third line is by Proposition 20 below.

Putting these bounds together yields the statement of the theorem.  $\blacksquare$

The following result is essentially a version of the one-shot decoupling proof of entanglement-distillation and random quantum coding, adapted so that the error is composed of a smoothing and a random coding component; its proof can be found in the appendix. Note that it gives an essentially matching lower bound to the upper bound we used in the proof of Theorem 2. It allows us to assess one of the max-entropy terms we encountered there in a new light.

**Proposition 20 (Cf. Buscemi/Datta [8] & Datta/Hsieh [12])** *Let  $U : A' \hookrightarrow B \otimes E$  be the Stinespring dilation of a quantum channel  $\mathcal{N}$  and  $|\phi\rangle \in AA'$  a state vector,  $|\psi\rangle := (\mathbb{1} \otimes U)|\phi\rangle \in ABC$ . Then, given  $\eta \geq 0$  and  $\epsilon > 0$ , there exists an entanglement-generating code for  $\mathcal{N}$ , creating a maximally entangled state of rank  $d$  with error  $\leq \eta + \epsilon$ , where*

$$d = \left\lceil \exp \left( H_{\min}^\eta(A|E)_\psi - 4 \log \frac{1}{\epsilon} \right) \right\rceil.$$

**Remark** We gave the very precise form of the bounds above to emphasize that if the strong converse holds in its exponential form for  $\mathcal{M}$ , in the sense that for every error rate  $c > 0$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log N_E(n, 1 - 2^{-cn} | \mathcal{M}) \leq f(c),$$

with some non-decreasing continuous function  $f(c)$  of  $c$  such that  $f(0) = 0$ , then there exists a similar function  $g(c)$  such that for  $\mathcal{N}$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log N_E(n, 1 - 2^{-cn} | \mathcal{N}) \leq Q^{(1)}(\mathcal{N}) + g(c).$$

In other words, if the error of  $\mathcal{M}$  converges to 1 exponentially for positive rates, then the error of  $\mathcal{N}$  converges to 1 exponentially for rates exceeding  $Q^{(1)}(\mathcal{N})$ .  $\blacksquare$

**Remark** The type I channel constructed in the proof of Lemma 17 is such that the composition  $UV$  of the Stinespring dilations and of channel and degrading channel, actually map the input space  $A'$  isometrically into  $F \otimes \text{Sym}^2(E) \subset F \otimes E \otimes E'$ , so that  $X_F = \mathbb{1}$ .

Looking at Lemma 18, we see that the symmetric channel constructed there has a dilation  $W : G \hookrightarrow \text{Sym}^2(E) \subset E \otimes E'$ , which is a restriction at the input of the “universal” symmetric channel  $S : \mathcal{L}(\text{Sym}^2(E)) \rightarrow \mathcal{L}(E)$  with the trivial Stinespring dilation

$$\text{Sym}^2(E) \hookrightarrow E \otimes E'.$$

To prove a full strong converse for all degradable channels, by Theorem 19 it is thus enough to show the strong converse for the channels  $S$ , for arbitrarily large dimension  $|E|$ . More precisely,  $|E| = 2|A||B|$  is enough for all degradable channels with given input and output spaces  $A$  and  $B$ . ■

### VIII. A SEMIDEFINITE PROGRAMMING APPROACH TO THE MIN-ENTROPY OF MULTIPLY SYMMETRIC STATES

In the proof of Theorem 2 we came across a term  $-H'_{\max}(AF^n|E'^n)$ ,  $\epsilon'$  being larger than the coding error we want to analyze. Similarly, in the proof of Theorem 14 we had  $-H'_{\max}(F^n|E'^n X)$ .

In both cases, assuming w.l.o.g. that the channel  $\mathcal{N}$  is of type I (Lemma 17) and using Lemma 18, we may view both expressions as  $-H'_{\max}(G^n|E'^n) = H'_{\min}(G^n|E'^n)$ , with respect to an input-output joint state of a symmetric channel  $\mathcal{M}^{\otimes n}$ . Lemma 18 also informs us that  $\mathcal{M}$  (or a trivial modification of  $\mathcal{M}$ ) has a Stinespring dilation  $W : G \hookrightarrow \text{Sym}^2(E) \subset E \otimes E'$ ; in fact, w.l.o.g.  $G = \text{Sym}^2(E)$  but we will not use this.

Now, in the proofs of Theorems 2 and 14 we only made use of the fact that  $\mathcal{M}^{\otimes n}$  is symmetric with respect to exchanging the entire output with the entire environment system. This symmetry was enough to show that for  $\epsilon' < \frac{1}{\sqrt{2}}$  this term can be bounded by a constant; we also remarked that for larger  $\epsilon'$  this kind of argument cannot be applied.

However, it is obvious that the channel has much more structure, which we ought to exploit. Indeed, it is symmetric with respect to exchanging the output and environment systems of any subset of the  $n$  instances of  $\mathcal{M}$  while leaving the others in place, i.e. for any  $I \subset [n]$ ,

$$\text{SWAP}_{EE'}^{\otimes I} W^{\otimes n} = W^{\otimes n},$$

and so the joint state of input, output and environment,  $|\psi\rangle^{G^n E^n E'^n} = (\mathbb{1} \otimes W)^{\otimes n} |\phi\rangle^{G^n G'^n}$ , satisfies similarly

$$\begin{aligned} \text{SWAP}_{EE'}^{\otimes I} |\psi\rangle^{G^n E^n E'^n} &= |\psi\rangle^{G^n E^n E'^n} \\ &= |\psi\rangle^{G^n E^n E'^n} \text{SWAP}_{EE'}^{\otimes I}, \end{aligned} \quad (19)$$

for all subsets  $I$ .

The semidefinite programming (SDP) formulation for the smoothed min-entropy is given by (cf. [48])

$$\begin{aligned} 2^{-H'_{\min}(G^n|E'^n)} &= \min \text{Tr } \sigma^{E^n} \quad \text{s.t.} \\ \rho^{G^n E^n E'^n} &\geq 0, \quad \text{Tr } \rho \leq 1, \\ \text{Tr } \rho \psi &\geq 1 - \epsilon'^2 =: \delta, \\ \rho^{G^n E^n} &\leq \mathbb{1}^{G^n} \otimes \sigma^{E^n}. \end{aligned}$$

By duality theory (cf. [48]) this value is equal to the dual SDP, given by

$$\begin{aligned} 2^{-H'_{\min}(G^n|E'^n)} &= \max \delta r - s \quad \text{s.t.} \\ r, s &\geq 0, \quad X^{G^n E^n} \geq 0, \\ r \psi^{G^n E^n E'^n} &\leq X^{G^n E^n} \otimes \mathbb{1}^{E'^n} + s \mathbb{1}, \\ \text{Tr }_{G^n} X &\leq \mathbb{1}^{E^n}. \end{aligned}$$

Note that we get an upper bound on  $H'_{\min}(G^n|E'^n)$  from every dual feasible point (a triple  $r, s, X$ ). The problem is to construct such a dual feasible point for each pure state  $\psi^{G^n E^n E'^n}$  with the symmetries (19) and each  $\delta > 0$ , such that  $\delta r - s \geq 2^{-\Omega(\sqrt{n})}$ . Since so far we were unable to find such a construction, we leave the problem at this point to the attention of the reader.

### IX. CONCLUSION

For degradable quantum channels, whose quantum and private capacities are known to be given by the single-letter maximization of the coherent information (which is then also additive on the class of all degradable channels), we have shown how to use the powerful min- and max-entropy calculus to derive bounds on the optimal quantum and private classical rate, for every finite blocklength  $n$ . These bounds improve on the well-known weak converse in that they give asymptotically the capacity as soon as the error (parametrized by the purified distance) is small enough: for  $Q$  this was  $\frac{1}{\sqrt{2}}$ , the error of a 50%-50% erasure channel, for  $P$  we could get  $\frac{1}{3\sqrt{2}}$ . Since this says equivalently that the minimum attainable error jumps from 0 to at least some threshold as the coding rate increases above the capacity, we speak of a “pretty strong” converse (halfway between a weak and a proper strong converse).

We have shown furthermore that it is enough to prove a strong converse for certain universal symmetric (degradable and anti-degradable) channels, namely those whose Stinespring dilation is the embedding of  $\text{Sym}^2(E)$  into  $E \otimes E'$  as a subspace; then the strong converse would follow for all degradable channels. To deal with these symmetric channels, and more generally with states exhibiting  $n$ -fold exchange symmetry between output and environment systems, we discussed briefly a semidefinite programming (SDP) approach. The viability of this approach stems from the fact that bounding the relevant min-entropy can be cast as a dual SDP, and so upper bounds may be obtained by any single dual feasible point. We have not been able to carry this part of the programme through yet.

Note that the proofs use the quantum AEP, but this does not mean that these results are restricted to i.i.d. channels. In fact, by using a standard discretization argument one can prove that for an arbitrary non-stationary memoryless channel  $\mathcal{N}_1 \otimes \dots \otimes \mathcal{N}_n$ , where each  $\mathcal{N}_t : \mathcal{L}(A) \rightarrow \mathcal{L}(B)$  is degradable, and sufficiently small error, the obviously defined  $\log N(n, \epsilon)$ ,  $\log N_E(n, \epsilon)$  and  $\log M(n, \epsilon, \delta)$  are asymptotically  $\sum_{t=1}^n Q^{(1)}(\mathcal{N}_t) \pm o(n)$  — cf. [1] and [49] for analogous statements for classical and classical-quantum channels, respectively.

Most channels of course are not degradable (or anti-degradable). For practically all these others we do not have any approach to obtain a strong or even just a pretty strong converse. One might speculate that other channels with additive coherent information, hence with a single-letter capacity formula, are also amenable to our method. But already the very attractive-looking class of *conjugate degradable* channels [7] poses new difficulties.

A related but different question is whether the *symmetric side channel-assisted quantum capacity*  $Q_{ss}(\mathcal{N})$  [41], which has an additive single-letter formula, obeys a pretty strong converse. Note that since arbitrary symmetric side-channels are permitted, including arbitrarily large 50%-50% erasure channels, the strong converse cannot hold for this capacity, since even infinite rate is achievable with error  $\frac{1}{\sqrt{2}}$ . Our present techniques, requiring bounds on the various system dimensions of the channel, do not to apply, and we seem to need new ideas.

**Note on related work.** In [37], Sharma and Warsi show that one may formulate upper bounds on the fidelity of codes in terms of the rate and so-called generalized divergences. Their approach doesn't appear to be related to ours, but it is conceivable that it may lead to proofs of strong converses for certain channels' quantum capacity. This however seems to presuppose that channel parameters derived from these divergences have strong additivity properties, which can only hold for channels with additive coherent information.

More precisely, the upper bound on the fidelity contained in [37, Thm. 1] is of no direct use, much as the trivial first steps in the proofs of our Theorems 2 and 14. The reason is that the bound explicitly depends on the code, via the joint input-output state. The only hope at this point is to control the maximum of said bound over all such input-output states. It is natural to expect that a first step might be to show that the maximum is attained on product states, but even that would leave nontrivial work to do. Crucially, the nature of the maximum bound is not addressed in [37]. Instead it is shown for the quantum erasure channel, that the bound, evaluated on the input-output state corresponding to maximally mixed input (which is indeed a tensor power), decreases exponentially.

This is the meaning of [37, Thm. 3], as one can discover from the calculation following its statement. Literally however, it says "The strong converse holds for the quantum erasure channel for the maximally entangled channel inputs", which might lead an unsuspecting reader to believe that indeed the strong converse is proved here, albeit perhaps with some restriction that is left vague. The concluding paragraph unfortunately repeats this claim in the stronger words "To summarize our results, we have given an exponential upper bound on the reliability of quantum information transmission", and "We then apply our bound to yield the first known example for exponential decay of reliability at rates above the capacity for quantum information transmission". Nothing could be further from the truth; not a single instance of exponential decay of fidelity above the capacity has been shown within the approach of [37], because the dependence on  $n$  of the maximum bound in [37, Thm. 1] is not generally understood. Indeed, claims such as the ones quoted above, would necessarily have to

involve a bound on all conceivable quantum codes, for large  $n$ , which seems difficult, to say the least. But the only code that [37, Thm. 3] covers is the trivial one of using the entire input bandwidth, not encoding at all. To analyze it, however, one hardly needs the machinery developed in [37]; it is an elementary exercise to show that *every* noisy channel exhibits exponential decay of fidelity for this code.

#### ACKNOWLEDGMENTS

We thank Mario Berta and Marco Tomamichel for discussions on strong converses in the context of quantum data compression with side information, Robert König and Stephanie Wehner for illuminating comments on strong converses, and Renato Renner and Frédéric Dupuis for sharing with us many of their insights regarding min-, max- and other entropies. In particular, we gratefully acknowledge Frédéric Dupuis' permission to use his result on the comparison between smooth max- and min-entropy (Lemma 9); and Robert König's suggestion of the name "pretty strong converse", as well as the PPT example in Section III. Normand Beaudry and Mark Wilde kindly pointed out several errors in a previous version of this preprint.

#### APPENDIX

Here we present the proofs of several auxiliary results used in the proof of the main result, which would have broken the flow of the text.

*Proof of Lemma 10:* Define the auxiliary state

$$\bar{\rho}^{ABX} := \sum_i p_i \rho_i^{AB} \otimes |i\rangle\langle i|^X,$$

so that the average of the  $\rho_i$  becomes  $\bar{\rho}^{AB} = \text{Tr}_X \bar{\rho}^{ABX}$ . Choosing purifications  $\psi_i^{ABC}$ , we can consider the following purification of  $\bar{\rho}^{ABX}$ :

$$|\varphi\rangle^{ABXCXY} = \sum_i \sqrt{p_i} |\psi_i\rangle^{ABC} |i\rangle^X |i\rangle^Y.$$

Then, using monotonicity (Lemma 6) and duality,

$$\begin{aligned} H_{\max}^\epsilon(A|B)_{\bar{\rho}^{AB}} &\geq H_{\max}^\epsilon(A|BX)_{\bar{\rho}^{ABX}} \\ &= -H_{\min}^\epsilon(A|CY)_\varphi, \end{aligned} \quad (20)$$

observing  $\varphi^{ACY} = \sum_i p_i \psi_i^{AC} \otimes |i\rangle\langle i|^Y$ .

Now, by definition of the smooth min-entropy, its exponential is give by the following optimization:

$$\begin{aligned} \Phi_\epsilon(\varphi^{A:CY}) &:= 2^{-H_{\min}^\epsilon(A|CY)_\varphi} \\ &= \min \text{Tr } \sigma^{CY} \text{ s.t.} \\ &\quad \tilde{\rho}^{ACY} \leq \mathbb{1}^A \otimes \sigma^{CY}, \\ &\quad \tilde{\rho} \geq 0, \text{Tr } \tilde{\rho} \leq 1, \\ &\quad F(\varphi^{ACY}, \tilde{\rho}) = \|\sqrt{\varphi} \sqrt{\tilde{\rho}}\|_1 \geq \sqrt{1 - \epsilon^2}. \end{aligned}$$

Since  $\varphi^{ACY}$  is invariant under phase unitaries on  $Y$ , we may assume w.l.o.g. that both  $\tilde{\rho}$  and  $\sigma$  have the same property, i.e. they may be assumed to be classical on  $Y$ :

$$\begin{aligned} \tilde{\rho}^{ACY} &= \sum_i q_i \tilde{\rho}_i^{AC} \otimes |i\rangle\langle i|^Y, \\ \sigma^{CY} &= \sum_i q_i \sigma_i^C \otimes |i\rangle\langle i|, \end{aligned}$$

where  $q_i \geq 0$ ,  $\sum_i q_i = 1$  and  $\tilde{\rho}_i \in \mathcal{S}_{\leq}(AC)$ ; furthermore  $\sigma_i \geq 0$ . With these notations, the objective function in the above optimization is  $\text{Tr } \sigma^{CY} = \sum_i q_i \text{Tr } \sigma_i^C$ , the first constraint is equivalent to  $\tilde{\rho}_i^{AC} \leq \mathbb{1}^A \otimes \sigma_i^C$  for all  $i$ , and

$$F(\varphi^{ACY}, \tilde{\rho}) = \sum_i \sqrt{p_i q_i} F(\psi_i^{AC}, \tilde{\rho}_i^{AC}).$$

Thus, observing that the  $\psi_i^{AC}$  are related to  $\psi_1^{AC} = \text{Tr}_B \psi_1$  by local unitaries, we have

$$\begin{aligned} \Phi_\epsilon(\varphi^{ACY}) &= \min \sum_i q_i \Phi_{\epsilon_i}(\psi_i^{AC}) \text{ s.t.} \\ &\quad \sum_i \sqrt{p_i q_i} \sqrt{1 - \epsilon_i^2} \geq \sqrt{1 - \epsilon^2} \\ &= \min \sum_i q_i \Phi_{\epsilon_i}(\psi_1^{AC}) \text{ s.t.} \\ &\quad \sum_i \sqrt{p_i q_i} \sqrt{1 - \epsilon_i^2} \geq \sqrt{1 - \epsilon^2}, \end{aligned}$$

where the variables are  $q_i$  and  $\epsilon_i$ .

Now, Cauchy-Schwarz inequality says

$$\sum_i \sqrt{p_i q_i} \sqrt{1 - \epsilon_i^2} \leq \sqrt{\sum_i p_i \sqrt{1 - \epsilon_i^2}} \sqrt{\sum_i q_i \sqrt{1 - \epsilon_i^2}}.$$

Hence the constraint implies that  $\sum_i q_i \sqrt{1 - \epsilon_i^2} \geq 1 - \epsilon^2$  and we get

$$\begin{aligned} \Phi_\epsilon(\varphi^{ACY}) &\geq \min \sum_i q_i \Phi_{\epsilon_i}(\psi_1^{AC}) \text{ s.t.} \\ &\quad \sum_i q_i \sqrt{1 - \epsilon_i^2} \geq 1 - \epsilon^2. \end{aligned}$$

For each  $i$ ,  $\Phi_{\epsilon_i}(\psi_1^{AC}) = \text{Tr } \sigma_i^C$  with  $0 \leq \tilde{\rho}_i \leq \mathbb{1}^A \otimes \sigma_i^C$ ,  $\text{Tr } \tilde{\rho}_i \leq 1$ , and  $F(\psi_1^{AC}, \tilde{\rho}_i) \geq \sqrt{1 - \epsilon_i^2}$ . Thus, forming  $\tilde{\omega} := \sum_i q_i \tilde{\rho}_i \in \mathcal{S}(AC)$  and  $\tilde{\sigma} = \sum_i q_i \sigma_i \geq 0$ , we have  $\text{Tr } \tilde{\omega} \leq 1$ ,  $\tilde{\omega} \leq \mathbb{1} \otimes \tilde{\sigma}$  and

$$F(\psi_1^{AC}, \tilde{\omega}) \geq \sum_i q_i \sqrt{1 - \epsilon_i^2} \geq 1 - \epsilon^2 =: \sqrt{1 - \hat{\epsilon}^2},$$

where  $\hat{\epsilon} \leq \epsilon \sqrt{2}$ .

This gives eventually

$$\Phi_\epsilon(\varphi^{ACY}) \geq \Phi_{\hat{\epsilon}}(\psi_1^{AC}),$$

so going back to Eq. (20), we arrive at

$$\begin{aligned} H_{\max}^\epsilon(A|B)_{\bar{\rho}} &\geq -H_{\min}^\epsilon(A|CY)_\varphi \\ &\geq -H_{\min}^{\hat{\epsilon}}(A|C)_{\psi_1} \\ &= H_{\max}^{\hat{\epsilon}}(A|C)_\rho \\ &\geq H_{\max}^{\epsilon\sqrt{2}}(A|C)_\rho, \end{aligned}$$

and we are done.  $\blacksquare$

*Proof of Lemma 11:* Fix purifications  $\psi_i^{ABC}$  of the  $\rho_i$ , so that  $\bar{\rho}$  can be purified as

$$|\psi\rangle^{ABCC_0} = \sum_{i=1}^M \sqrt{p_i} |\psi_i\rangle^{ABC} |i\rangle^{C_0}.$$

We use the following characterization of smooth max-entropies (cf. [45]):

$$\begin{aligned} 2^{H_{\max}^\epsilon(A|B)_{\rho_i}} &= \min \|\text{Tr}_A Z_i\| \text{ s.t.} \\ F(\psi_i, \psi'_i) &\geq \sqrt{1 - \epsilon^2}, \\ \psi'_i &\leq Z_i^{AB} \otimes \mathbb{1}^C. \end{aligned}$$

Fix optimal  $|\psi'_i\rangle \in ABC$ , such that  $\langle \psi_i | \psi'_i \rangle = F(\psi_i, \psi'_i) \geq \sqrt{1 - \epsilon^2}$ , and  $Z_i \geq 0$ . Let  $\lambda = \max_i \|\text{Tr}_A Z_i\|$  and define

$$|\psi'\rangle^{ABCC_0} := \sum_{i=1}^M \sqrt{p_i} |\psi'_i\rangle^{ABC} |i\rangle^{C_0},$$

so that

$$F(\psi, \psi') = \langle \psi | \psi' \rangle = \sum_i p_i \langle \psi_i | \psi'_i \rangle \geq \sqrt{1 - \epsilon^2}.$$

Furthermore, using Hayashi's pinching inequality [23], [28] in the second line,

$$\begin{aligned} |\psi'\rangle\langle\psi'| &= \sum_{ij=1}^M \sqrt{p_i p_j} |\psi'_i\rangle\langle\psi'_j| \otimes |i\rangle\langle j| \\ &\leq M \sum_{i=1}^M p_i \psi_i'^{ABC} \otimes |i\rangle\langle i|^{C_0} \\ &\leq \sum_i M p_i Z_i^{AB} \otimes \mathbb{1}^C \otimes \mathbb{1}^{C_0} \\ &=: Z^{AB} \otimes \mathbb{1}^{CC_0}. \end{aligned}$$

I.e.,  $\psi'$  and  $Z$  are feasible for  $\bar{\rho}$ , and the objective function value

$$\begin{aligned} \|\text{Tr}_A Z\| &= \left\| \sum_i M p_i \text{Tr}_A Z_i \right\| \\ &\leq \sum_i M p_i \|\text{Tr}_A Z_i\| \\ &\leq M \lambda \end{aligned}$$

gives an upper bound to  $2^{H_{\max}^\epsilon(A|B)_{\bar{\rho}}}$ . Thus we can conclude

$$\begin{aligned} H_{\max}^\epsilon(A|B)_{\bar{\rho}} &\leq \log \lambda + \log M \\ &= \max_i H_{\max}^\epsilon(A|B)_{\rho_i} + \log M, \end{aligned}$$

as advertised.  $\blacksquare$

*Proof of Proposition 13:* To get bounds valid for all  $n$ , we use well-known tail estimates for sums of independent random variables due to Hoeffding [14]. To wit, consider the discrete random variable  $X$  with minimum non-zero probability  $\min_x P_X(x) =: 2^{-\mu}$  and let  $L = L(X) := -\log P_X(X)$ , such that  $0 \leq L \leq \mu$  with probability 1, and  $\mathbb{E}L = H(P)$ . Then, for i.i.d. realizations  $X_1, X_2, \dots, X_n$  of  $X$ , and associated  $L_i$ , Hoeffding's inequality states

$$\begin{aligned} \Pr \left\{ \sum_{i=1}^n L_i > nH(P) + \Delta \sqrt{n} \right\} &\leq e^{-\frac{2\Delta^2}{\mu^2}}, \\ \Pr \left\{ \sum_{i=1}^n L_i < nH(P) - \Delta \sqrt{n} \right\} &\leq e^{-\frac{2\Delta^2}{\mu^2}}. \end{aligned} \quad (21)$$

We can use these bounds to construct typical projectors for a state  $\rho^{\otimes n}$ ,  $\rho \in \mathcal{S}(\mathcal{H})$ , in the usual way. Let  $\rho = \sum_x \lambda_x |x\rangle\langle x|$

be a diagonalization, so that  $\lambda_x$  can be interpreted as a probability distribution on the  $x$ . Define two projectors

$$P_{\rho^{\otimes n}}^{+\Delta} := \sum_{x^n \in \mathcal{T}_{\lambda^{\otimes n}}^{+\Delta}} |x^n\rangle\langle x^n| \text{ with}$$

$$\mathcal{T}_{\lambda^{\otimes n}}^{+\Delta} := \left\{ x^n = x_1 \dots x_n : \sum_i -\log \lambda_{x_i} \leq nS(\rho) + \Delta\sqrt{n} \right\},$$

and

$$P_{\rho^{\otimes n}}^{-\Delta} := \sum_{x^n \in \mathcal{T}_{\lambda^{\otimes n}}^{-\Delta}} |x^n\rangle\langle x^n| \text{ with}$$

$$\mathcal{T}_{\lambda^{\otimes n}}^{-\Delta} := \left\{ x^n = x_1 \dots x_n : \sum_i -\log \lambda_{x_i} \geq nS(\rho) - \Delta\sqrt{n} \right\}.$$

By Eq. (21),

$$\begin{aligned} \text{Tr } \rho^{\otimes n} P_{\rho^{\otimes n}}^{+\Delta} &\geq 1 - e^{-\frac{2\Delta^2}{\mu^2}}, \\ \text{Tr } \rho^{\otimes n} P_{\rho^{\otimes n}}^{-\Delta} &\geq 1 - e^{-\frac{2\Delta^2}{\mu^2}}, \end{aligned}$$

where  $\mu = \log \|\rho^{-1}\|$ .

Now, for a pure tripartite state  $|\psi\rangle \in ABC$ , let  $\Delta > 0$  and consider the projectors

$$\begin{aligned} P_B^+ &:= P_{\rho_B^{\otimes n}}^{+\Delta\mu_B}, \\ P_C^- &:= P_{\rho_C^{\otimes n}}^{-\Delta\mu_C}. \end{aligned}$$

Defining  $|\Psi'\rangle := (\mathbb{1}_A \otimes P_B^+ \otimes P_C^-)|\psi\rangle^{\otimes n}$ , clearly we have

$$\begin{aligned} \langle \Psi' | \psi \rangle^{\otimes n} &= \langle \psi |^{\otimes n} (\mathbb{1}_A \otimes P_B^+ \otimes P_C^-) | \psi \rangle^{\otimes n} \\ &\geq 1 - 2e^{-2\Delta^2} \\ &\stackrel{!}{\geq} \sqrt{1 - \epsilon^2}, \end{aligned}$$

for  $\Delta = \sqrt{\ln \frac{2}{\epsilon}}$ . By definition

$$\Psi'^{C^n} \simeq \Psi'^{A^n B^n} \leq 2^{-nS(AB) + \Delta\mu_C\sqrt{n}} (\mathbb{1}^{A^n} \otimes P_B^+).$$

On the other hand, we just need to rescale  $P_B^+$  by its trace,  $\sigma := \frac{1}{\text{Tr } P_B^+} P_B^+$  to get an eligible state in the definition of  $H_{\min}^\epsilon(A|B)$ . Note that  $\text{Tr } P_B^+ \leq 2^{nS(\rho) + \Delta\mu_B\sqrt{n}}$ , hence

$$\Psi'^{A^n B^n} \leq 2^{-nS(A|B) + \Delta(\mu_B + \mu_C)\sqrt{n}} (\mathbb{1}^{A^n} \otimes \sigma^{B^n}),$$

thus showing

$$H_{\min}^\epsilon(A|B) \geq nS(A|B) - (\mu_B + \mu_C)\sqrt{n \ln \frac{2}{\epsilon}}.$$

The upper bound on  $H_{\max}^\epsilon(A|B)$  follows by the duality of the min- and max-entropies, as well as that of the conditional von Neumann entropy:  $S(A|B) = -S(A|C)$ . ■

*Proof of Proposition 20:* For a  $d$ -dimensional projector  $Q$  on  $A$ , write

$$\sqrt{\frac{|A|}{d}} (Q \otimes \mathbb{1}) |\psi^{ABE}\rangle =: \sqrt{t_Q} |\tilde{\psi}_Q\rangle^{ABE},$$

where  $\sqrt{t_Q}$  is the normalisation of the left hand side and  $|\tilde{\psi}_Q\rangle^{ABE}$  is a state. Our goal is to show that we can find

$Q$  such that  $\tilde{\psi}_Q^{AE}$  is close to a product state. To be precise, the claim is that there exists  $\varphi \in \mathcal{S}_{\leq}(E)$  and  $Q$  such that

$$P(\tilde{\psi}_Q, \tau_Q \otimes \varphi^E) \leq \eta + 2^{-\frac{1}{4}(H_{\min}^\eta(A|E)_\psi - \log d)}. \quad (22)$$

Then, using the familiar decoupling argument, there is a cptp map  $\mathcal{D}$  acting on  $B$  such that

$$P((\text{id} \otimes \mathcal{D})\tilde{\psi}_Q^A B, \Phi_{QQ'}) \leq \eta + 2^{-\frac{1}{4}(H_{\min}^\eta(A|E)_\psi - \log d)},$$

where  $\Phi_{QQ'}$  is a maximally entangled state. Choosing

$$|\tilde{\phi}_Q\rangle^{AA'} := \sqrt{|A|} dt_Q (Q \otimes \mathbb{1}) |\phi\rangle$$

as the input state, so that  $|\tilde{\psi}_Q\rangle^{ABE} = (\mathbb{1} \otimes U) |\tilde{\phi}_Q\rangle$ , completes the entanglement-generating code. Choosing  $\log d \leq H_{\min}^\eta(A|E)_\psi - 4 \log \frac{1}{\epsilon}$  guarantees that its error is  $\leq \eta + \epsilon$ .

To prove Eq. (22), choose a  $\varphi \in \mathcal{S}_{\leq}(ABE)$  with  $P(\varphi, \psi) \leq \eta$  and  $H_{\min}^\eta(A|E)_\psi = H_{\min}(A|E)_\varphi$ . Consider the cptp map

$$\mathcal{P} : \rho \mapsto \int dQ \frac{|A|}{d} Q \rho Q^\dagger \otimes |Q\rangle\langle Q|,$$

where  $|Q\rangle$  are orthogonal labels of a dummy system. By the contractiveness of the purified distance, we have

$$P((\mathcal{P} \otimes \text{id})\varphi^{AE}, (\mathcal{P} \otimes \text{id})\psi^{AE}) \leq \eta. \quad (23)$$

We also have  $\int dQ t_Q = 1$ .

Now, Lemma 21 below tells us

$$\|(\mathcal{P} \otimes \text{id})\varphi^{AE} - (\mathcal{P} \otimes \text{id})(\tau_A \otimes \varphi^E)\|_1 \leq 2^{-\frac{1}{2}(H_{\min}(A|E)_\varphi - \log d)},$$

noting

$$(\mathcal{P} \otimes \text{id})(\tau_A \otimes \varphi^E) = \int dQ \tau_Q \otimes \varphi^E \otimes |Q\rangle\langle Q|,$$

and that the trace norm on the left hand side is

$$\int Q \left\| \frac{|A|}{d} (Q \otimes \mathbb{1}) \varphi^{AE} (Q \otimes \mathbb{1})^\dagger - \tau_Q \otimes \varphi^E \right\|_1.$$

By Eq. (1), the trace norm bound implies

$$P((\mathcal{P} \otimes \text{id})\varphi^{AE}, (\mathcal{P} \otimes \text{id})(\tau_A \otimes \varphi^E)) \leq 2^{-\frac{1}{4}(H_{\min}(A|E)_\varphi - \log d)}.$$

Substituting  $H_{\min}(A|E)_\varphi = H_{\min}^\eta(A|E)_\psi$  and using Eq. (23) with the triangle inequality for the purified distance, we get

$$\begin{aligned} P((\mathcal{P} \otimes \text{id})\psi^{AE}, (\mathcal{P} \otimes \text{id})(\tau_A \otimes \varphi^E)) \\ \leq \eta + 2^{-\frac{1}{4}(H_{\min}^\eta(A|E)_\psi - \log d)} =: \delta. \end{aligned}$$

Equivalently, inserting the definition of  $\tilde{\psi}_Q$  and  $t_Q$ :

$$\begin{aligned} \sqrt{1 - \delta^2} &\leq F((\mathcal{P} \otimes \text{id})\psi^{AE}, (\mathcal{P} \otimes \text{id})(\tau_A \otimes \varphi^E)) \\ &= \int dQ \left\| \sqrt{t_Q} \tilde{\psi}_Q^{AE} \sqrt{\tau_Q \otimes \varphi^E} \right\|_1 \\ &= \int dQ \sqrt{t_Q} F(\tilde{\psi}_Q^{AE}, \tau_Q \otimes \varphi^E). \end{aligned}$$

Since finally, by the concavity of the square root,

$$\int dQ \sqrt{t_Q} \leq \sqrt{\int dQ t_Q} = 1,$$

this implies that there exists  $Q$  in the previous integral with  $F(\tilde{\psi}_Q^{AE}, \tau_Q \otimes \varphi^E) \geq \sqrt{1 - \delta^2}$ , which is precisely Eq. (22). ■

**Lemma 21 (Berta [5])** *Let  $|\varphi\rangle \in ABC$  be a state vector. Picking a  $d$ -dimensional projector  $Q$  uniformly (i.e. from the unitarily invariant measure  $dQ$ ), we have*

$$\int dQ \left\| \frac{|A|}{d} (Q \otimes \mathbb{I}) \varphi^{AE} (Q \otimes \mathbb{I})^\dagger - \tau_Q \otimes \psi^E \right\|_1 \leq 2^{-\frac{1}{2}(H_{\min}(A|E) - \log d)},$$

with the maximally mixed state  $\tau_Q = \frac{1}{d}Q \in \mathcal{S}(A)$  on the support of  $Q$ . ■

## REFERENCES

- [1] R. Ahlswede. Beiträge zur Shannonschen Informationstheorie im Falle nichtstationärer Kanäle. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **10** (1968), 1–42.
- [2] R. Ahlswede. Elimination of correlation in random codes for arbitrarily varying channels. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **44** (1978), 159–175.
- [3] K. Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical J.*, **19** (1967), 357–367.
- [4] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor and A. Winter. The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels. arXiv:0912.5537 (2009).
- [5] M. Berta. *Single-shot Quantum State Merging*. Diploma thesis, ETH Zürich (2009); arXiv:0912.4495.
- [6] M. Berta, M. Christandl and R. Renner. The Quantum Reverse Shannon Theorem based on One-Shot Information Theory. *Commun. Math. Phys.*, **306** (2011), 579–615.
- [7] K. Brádler, N. Dutil, P. Hayden and A. Muhammad. Conjugate degradability and the quantum capacity of cloning channels. *J. Math. Phys.*, **51** (2010), 072201.
- [8] F. Buscemi and N. Datta. The quantum capacity of channels with arbitrarily correlated noise. *IEEE Trans. Inf. Theory*, **56** (2010), 1447–1460.
- [9] N. Cai, A. Winter and R. W. Yeung. Quantum Privacy and Quantum Wiretap Channels. *Problems Inf. Transm.*, **40**, 4 (2004), 318–336.
- [10] M. Christandl, R. König and R. Renner. Postselection Technique for Quantum Channels with Applications to Quantum Cryptography. *Phys. Rev. Lett.*, **102** (2009), 020504.
- [11] T. Cubitt, M.-B. Ruskai and G. Smith. The structure of degradable quantum channels. *J. Math. Phys.*, **49** (2008), 102104.
- [12] N. Datta and M.-H. Hsieh. The apex of the family tree of protocols: Optimal rates and resource inequalities. *New J. Phys.*, **13** (2011), 093042; arXiv:1103.1135.
- [13] N. Datta, M. Mosonyi, M.-H. Hsieh and F. Brandão. Strong converses for classical information transmission and hypothesis testing. arXiv:1106.3089 (2011).
- [14] A. Dembo and O. Zeitouni. *Large Deviations: Techniques and Applications*, 2nd edition. Applications of Mathematics, vol. 38, Springer Verlag, Berlin Heidelberg, 1998.
- [15] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Inf. Theory*, **51** (2005), 44–55.
- [16] I. Devetak and P. W. Shor. The capacity of a quantum channel for simultaneous transmission of classical and quantum information. *Comm. Math. Phys.*, **256** 2005, 287–303.
- [17] D. P. DiVincenzo, P. W. Shor and J. A. Smolin. Quantum-channel capacity of very noisy channels. *Phys. Rev. A*, **57**, 2 (1998), 830–839.
- [18] T. Dorlas and C. Morgan. The invalidity of a strong capacity for a quantum channel with memory. *Phys. Rev. A*, **84** (2011), 042318; arXiv:1108.4282.
- [19] F. Dupuis. *The decoupling approach to quantum information theory*. PhD thesis, Université de Montréal (2009); arXiv:1004.1641.
- [20] F. Dupuis. private communication (January 2013).
- [21] F. Dupuis, M. Berta, J. Wullschlegel and R. Renner. The decoupling theorem. arXiv:1012.6044 (2010).
- [22] C. A. Fuchs and J. van de Graaf. Cryptographic Distinguishability Measures for Quantum-Mechanical States. *IEEE Trans. Inf. Theory*, **45** (1999), 1216–1227.
- [23] M. Hayashi. Optimal sequence of POVMs in the sense of Steins Lemma in quantum hypothesis testing. *J. Phys. A: Math. Gen.*, **35**, 5 (2002), 10759–10773; arXiv:quant-ph/0107004.
- [24] R. Jozsa. Fidelity for mixed quantum states. *J. Mod. Opt.*, **41** (1994), 2315–2323.
- [25] R. König and S. Wehner. A strong converse for classical channel coding using entangled inputs. *Phys. Rev. Lett.*, **103** (2009), 070504.
- [26] D. Kretschmann and R. F. Werner. Tema con variazioni: quantum channel capacity. *New J. Phys.*, **6** (2004), 6.
- [27] S. Lloyd. Capacity of the noisy quantum channel. *Phys. Rev. A*, **55** (1997), 1613–1622.
- [28] T. Ogawa and M. Hayashi. On Error Exponents in Quantum Hypothesis Testing. *IEEE Trans. Inf. Theory*, **50** (2004), 1368–1372; arXiv:quant-ph/0110125.
- [29] T. Ogawa and H. Nagaoka. Strong converse to the quantum channel coding theorem. *IEEE Trans. Inf. Theory*, **45** (1999), 2486–2489.
- [30] E. M. Rains. A Semidefinite Program for Distillable Entanglement. *IEEE Trans. Inf. Theory*, **47** (2001), 2921–2933.
- [31] J. M. Renes and R. Renner. Noisy channel coding via privacy amplification and information reconciliation. *IEEE Trans. Inf. Theory*, **57** (2011), 7377–7385.
- [32] R. Renner. *Security of Quantum Key Distribution*. PhD thesis, ETH Zürich (2005); arXiv:quant-ph/0512258.
- [33] R. Renner. Symmetry of large physical systems implies independence of subsystems. *Nature Physics*, **3** (2007), 645–649.
- [34] B. Schumacher. Sending entanglement through noisy quantum channels. *Phys. Rev. A*, **54** (1996), 2614–2628.
- [35] B. Schumacher and M. A. Nielsen. Quantum data processing and error correction. *Phys. Rev. A*, **54** (1996), 2629–2635.
- [36] C. E. Shannon. A Mathematical Theory of Communication. *Bell Syst. Tech. J.*, **27** (1948), 379–423 & 623–656.
- [37] N. Sharma and N. A. Warsi. Fundamental bound on the reliability of quantum information transmission. *Phys. Rev. Lett.*, **110** (2013), 080501.
- [38] P. W. Shor. The quantum channel capacity and coherent information. Lecture notes, MSRI Workshop on Quantum Computation, San Francisco, November 2002.
- [39] P. W. Shor. Capacities of quantum channels and how to find them. *Math. Program. Ser. B*, **97** (2003), 311–335.
- [40] P. W. Shor and J. A. Smolin. Quantum Error-Correcting Codes Need Not Completely Reveal the Error Syndrome. arXiv:quant-ph/9604006 (1996).
- [41] G. Smith, J. A. Smolin and A. Winter. The Quantum Capacity With Symmetric Side Channels. *IEEE Trans. Inf. Theory*, **54** (2008), 4208–4217.
- [42] G. Smith, J. M. Renes and J. A. Smolin. Structured Codes Improve the Bennett-Brassard-84 Quantum Key Rate. *Phys. Rev. Lett.*, **100** (2008), 170502.
- [43] G. Smith. The private classical capacity with a symmetric side channel and its application to quantum cryptography. *Phys. Rev. A*, **78** (2008), 022306.
- [44] G. Smith and J. Yard. Quantum communication with zero-capacity channels. *Science*, **321** (2008), 1812–1815; arXiv:0807.4935.
- [45] M. Tomamichel. *A framework for non-asymptotic quantum information theory*. PhD thesis, Department of Physics, ETH Zürich (2011). arXiv:1203.2142.
- [46] M. Tomamichel, R. Colbeck, and R. Renner. Duality between smooth min- and max-entropies. *IEEE Trans. Inf. Theory*, **56** (2010), 4674–4681.
- [47] A. Uhlmann. The “Transition Probability” in the State Space of a \*-Algebra. *Rep. Math. Phys.*, **9** (1976), 273–279.
- [48] A. Vitanov, F. Dupuis, M. Tomamichel and R. Renner. Chain Rules for Smooth Min- and Max-Entropies. arXiv:1205.5231.
- [49] A. Winter. *Coding Theorems of Quantum Information Theory*. PhD thesis, Department of Mathematics, Universität Bielefeld (1999). arXiv:quant-ph/9907077.
- [50] A. Winter. Coding theorem and strong converse for quantum channels. *IEEE Trans. Inf. Theory*, **45** (1999), 2481–2485.